



MultiGML: Multimodal graph machine learning for prediction of adverse drug events

Sophia Krix ^{a,b,c,1}, Lauren Nicole DeLong ^{a,d,1}, Sumit Madan ^{a,e}, Daniel Domingo-Fernández ^{a,c,f}, Ashar Ahmad ^{b,g}, Sheraz Gul ^{h,i}, Andrea Zaliani ^{h,i}, Holger Fröhlich ^{a,b,*}

^a Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53757, Sankt Augustin, Germany

^b Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, 53115, Bonn, Germany

^c Fraunhofer Center for Machine Learning, Germany

^d Artificial Intelligence and its Applications Institute, School of Informatics, University of Edinburgh, 10 Crichton Street, EH8 9AB, UK

^e Department of Computer Science, University of Bonn, 53115, Bonn, Germany

^f Enveda Biosciences, Boulder, CO, 80301, USA

^g Grunenthal GmbH, 52099, Aachen, Germany

^h Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, Schnackenburgallee 114, 22525, Hamburg, Germany

ⁱ Fraunhofer Cluster of Excellence for Immune-Mediated Diseases CIMD, Schnackenburgallee 114, 22525, Hamburg, Germany

ARTICLE INFO

Keywords:

Machine learning
Knowledge graph
Adverse event
Graph neural network
Graph attention network
Graph convolutional network

ABSTRACT

Adverse drug events constitute a major challenge for the success of clinical trials. Several computational strategies have been suggested to estimate the risk of adverse drug events in preclinical drug development. While these approaches have demonstrated high utility in practice, they are at the same time limited to specific information sources. Thus, many current computational approaches neglect a wealth of information which results from the integration of different data sources, such as biological protein function, gene expression, chemical compound structure, cell-based imaging and others. In this work we propose an integrative and explainable multimodal Graph Machine Learning approach (MultiGML), which fuses knowledge graphs with multiple further data modalities to predict drug related adverse events and general drug target-phenotype associations. MultiGML demonstrates excellent prediction performance compared to alternative algorithms, including various traditional knowledge graph embedding techniques. MultiGML distinguishes itself from alternative techniques by providing in-depth explanations of model predictions, which point towards biological mechanisms associated with predictions of an adverse drug event. Hence, MultiGML could be a versatile tool to support decision making in preclinical drug development.

* Corresponding author. Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53757, Sankt Augustin, Germany.

E-mail address: holger.froehlich@scai.fraunhofer.de (H. Fröhlich).

¹ Shared first-authorship.

1. Introduction

Adverse drug events (ADEs) are defined as an injury resulting from the use of a drug, including harm caused by the drug (adverse drug reactions and overdoses) and harm from the use of the drug (including dose reductions and discontinuations of drug therapy) [1]. Noteworthy, the appearance of an ADE can be associated with the choice of the primary target protein or properties of the chemical structure of a drug. Experimental approaches to address potential ADEs (e.g. liver-toxicity) based on animal and tissue models are well established in pharma research. Yet, results obtained in such model systems may not always reflect the situation in humans, and there are ethical concerns regarding the use of animal models. Furthermore, reliable model systems do not exist for all indication areas. Computational approaches which model relevant aspects of human biology could bridge this gap and provide supportive information regarding potential ADE prediction. Hence, there is a strong interest in computational strategies. Computational ADE prediction has been tackled with the help of several data sources, such as human genetics [2–4], chemical structures [5–10], high-throughput literature mining [11], gene expression data [12,13], protein sequences [14], electronic health records [15] and data from electronic pharmacovigilance systems such as the FDA Adverse Event Reporting System [16]. While each of these approaches have their own merits, they also come along with unavoidable limitations: For example, genetic variants associated with a certain phenotype may not be identifiable in genome-wide association studies due to lack of statistical power. Chemical compound structure can inform about binding affinity to a given target, but does not cover the question whether the choice of a specific target should per se raise safety concerns due to the expected biological downstream consequences. Electronic health records can inform about real-world post-marketing aspects of drugs, but have limited utility in the preclinical drug development phase due to the lack of quantitative biological data.

An alternative strategy is to use biological networks, which represent a rich resource of relational information. In this context knowledge graphs (KGs) have become popular due to their ability to accurately represent multiple types of relationships between different entities [17–19]. That means KGs are multi-relational graphs with entities as nodes and their relations as edges. Relations are represented as triples of (source entity, relation type, target entity). KGs often incorporate a variety of heterogeneous information in the form of different node and edge types. In recent years, numerous knowledge graphs have been published such as OpenBioLink [20], Hetionet [21], PharmKG [22] or CTKG [23], a knowledge graph on clinical trials. These comprehensive KGs contain a variety of entity types and relation types which model biology as accurately as possible and can be applied to multiple tasks due to their versatile design.

From a network-perspective, ADE prediction can be formulated as a link prediction task in a KG, either between a compound and an unwanted phenotype, or between a drug target and a phenotype. Earlier approaches extracted manually crafted features of the topology of the KG by using the neighborhood information of each node [24] or by extracting local information indexes and path information [25]. Other authors used an enrichment test of known causes of ADEs to construct features that were subsequently employed in a machine learning algorithm [26]. Another network-based approach used structural information of the drug molecules for a logistic regression model [27]. As interactions between co-prescribed drugs are also a possible cause of ADEs [28], the prediction of drug-drug interactions has been the focus of several approaches. For that purpose, similarity measures [29] and representative KG embeddings of chemical drug structures via neural networks [30] have been used in prediction approaches. Other authors proposed network representation learning techniques and graph regularized matrix factorization for predicting ADEs of individual drugs [31,32]. Also, ensembles of several learning techniques have been tested [33].

From a methodological point of view, link prediction in KGs can be addressed by first learning an embedding of the graph structure in Euclidean space. Essentially, the KG embedding is a low-dimensional representation which captures key information about entities and their relations. Typically, entities with similar embeddings are also similar in the original space. Hence, we can assess the likelihood that two entities should be connected by a relation type. In addition to established network representation learning methods such as TransE [34], ComplEx [35], DistMult [36], RotatE [37], DeepWalk [38] and node2vec [39], graph neural networks (GNNs) have emerged as an efficient machine learning method. GNNs were first introduced by Scarselli et al. [40]. Subsequently, graph convolutional neural networks (GCNs) [41] and graph attention networks (GATs) [42] were developed as variants of GNNs. GNNs have been successfully applied to various tasks in network analytics, including clustering [43] and disease classification [44], prediction of molecular fingerprints [45] and protein interfaces [46], as well as prediction of drug-protein interactions [47] and poly-pharmacological side effects [48]. A GNN has also been used on a drug-disease graph for ADE prediction [49]. This approach has been further developed by combining two GNNs for graph and node embedding in a hybrid approach to predict ADEs via a matrix completion process [50]. Recently, a graph convolutional autoencoder approach coupled with an attention mechanism has been suggested, leveraging the pairwise attributes for drug-related ADE prediction in a heterogeneous graph [51].

A limitation of these existing KG focused approaches is that they neglect any orthogonal information, including genetic associations, chemical compound structure, gene expression signatures and cell morphology changes. The aim of this paper is thus to address limitations of previous work by developing a Graph Machine Learning approach, which integrates biological networks, genetic variant to phenotype association, gene expression, cell based imaging, protein sequence information, clinical concept embeddings as well as chemical compound fingerprints into one end-to-end trainable algorithm. The idea is thus to integrate a large number of potentially relevant sources of evidence to predict potential ADEs that could occur during clinical trials. Consequently, well-informed ADE predictions could reduce the risk of late and costly failures [52]. To do so, we built a dedicated KG and designed a novel GNN architecture tailored for ADE prediction. The KG consists of multi-relational and heterogeneous information collected from 14 different databases. The KG is enriched with multi-modal features for each node in order to capture various relevant biomedical data in addition to graph topology. As opposed to state-of-the-art approaches, our proposed MultiGML model is thus designed to integrate multi-modal and in particular also quantitative input data (e.g. gene expression). We demonstrate the superior prediction performance of our approach by

comparing it with several state-of-the-art models. Moreover, we introduce a technique to make model predictions explainable, which is crucial in the context of an application in the early phases of drug development. Based on a number of examples we show that our method in this way allows for pointing towards the biological mechanisms associated with a given ADE prediction. Finally, we provide literature evidence for some of the predictions made by our GNN method. The source code and the Python package of MultiGML is available on [GitHub](https://github.com/SCAI-BIO/MultiGML) (<https://github.com/SCAI-BIO/MultiGML>).

2. Results

2.1. Link prediction performance

In the following, we show the prediction performance of our MultiGML models compared to various competing methods for link prediction in KGs. We first evaluated MultiGML for the task of predicting any link in the KG and second for the more specific task of adverse drug event prediction. Regarding the task of general link prediction, our MultiGML-RGCN model reached a performance of 0.808 area under precision recall curve (AUPR), and the MultiGML-RGAT model reached an AUPR of 0.798, both outperforming all competing methods (Table 1) by at least $\sim 5\%$. These results show the superiority of our graph neural network-based architecture compared to more shallow knowledge graph embedding techniques. When using a randomly initialized vector embedding instead of the multi-modal feature embedding (i.e. essentially only learning from the graph topology), there is only a slight decrease in performance. That means, our MultiGML models already allowed us to reach a high prediction performance by the graph structure alone, which could be further enhanced by adding multi-modal node features.

We focused subsequently on the task of predicting links between compounds and ADEs. For that purpose we employed a version of our MultiGML model for which we specifically optimized hyperparameters on the validation set with respect to the loss for this specific relation type. Once again, all MultiGML variants performed better than all competing methods with AUROC and AUPR close to 1 (Table 2). TransE performed very poorly on the ADE prediction task, which could result from the limitations that this approach has to model complex relations, such as one-to-many, many-to-one, many-to-many, which can occur especially in the context of drugs and phenotypes. Performance gains were highly significant compared to the Random Forest approach by Wang et al. Notably, reported performance measures were based on the negative sampling scheme explained in section 3.3.1. When increasing the ratio of negative samples in the test set from 1:1 to 1000:1 AUROC and AUPR remained stable (see Suppl. Fig. 1).

Next, we evaluated the model performance for predicting links between genes and phenotypes, which would be of relevance in the context of target selection. For this purpose we used our MultiGML models which were trained for general link prediction. Once again all variants of the MultiGML model outperformed competing methods with AUROC ~ 0.89 and AUPR ~ 0.83 (Table 3). Even though these models were not optimized for the given task, they still achieved a high prediction performance which reflects that they are not trained to be biased towards any kind of relation type and advocates for a strong generalizability of the models.

As a further analysis, we explored which feature modalities contributed most to our model's predictions. Notably, in both MultiGML variants, available protein and drug features played an important role, i.e. were selected during the hyperparameter optimization (see Suppl. Fig. 2). More specifically, the molecular fingerprint of the drugs as well as the gene ontology fingerprint of the proteins were found to be the best choices of node features for the prediction of ADEs with our MultiGML models. Additionally, gene expression signatures of drugs were identified as relevant. When replacing these node features by randomly initialized vectors the performance of MultiGML variants did not suffer significantly, i.e. the graph topology contributed most of the relevant information. Despite this finding, we would like to point out that the inclusion of multimodal node features could enhance the interpretation of models, as shown later.

Altogether our results indicate that MultiGML demonstrates superior prediction performance compared to baseline methods for the prediction of adverse event prediction and general phenotypes.

Table 1

Model performance results for general relation prediction. The table shows the test results of several competing KG embedding methods, including TransE, RotatE, ComplEx, DistMult, DeepWalk and node2vec, as well as our two tested MultiGML model variants. Best results are marked in bold. Both RGCN and RGAT variants of the MultiGML model were tested with two types of input features. The model variant “multimodal” refers to the use of several modalities for each node type described in section 3.1.2. In the model variant “basic” all input features have been initialized with the Xavier-Glorot method, i.e. the model effectively learns from the topology only.

Model	AUROC	AUPR
TransE	0.667	0.633
RotatE	0.793	0.759
ComplEx	0.757	0.699
DistMult	0.765	0.696
DeepWalk	0.648	0.622
Node2Vec	0.807	0.794
MultiGML-RGCN (basic)	0.847	0.787
MultiGML-RGAT (basic)	0.843	0.793
MultiGML-RGCN (multimodal)	0.859	0.808
MultiGML-RGAT (multimodal)	0.845	0.798

Table 2

Model performance results for predicting a novel relation between a drug and an ADE. Test results of competing KG embedding methods, including TransE, RotatE, ComplEx, DistMult, DeepWalk, node2vec and additionally Random Forest for adverse drug event prediction in comparison to our MultiGML models. Best results are marked in bold. Both MultiGML-RGCN and -RGAT variants were tested with basic and multimodal input features. The model variant “multimodal” refers to the use of several modalities for each node type described in section 3.1.2. In the model variant “basic” all input features have been initialized with the Xavier-Glorot method, i.e. the model effectively learns from the topology only.

Model	AUROC	AUPR
TransE	0.293	0.389
RotatE	0.943	0.915
ComplEx	0.884	0.934
DistMult	0.963	0.966
DeepWalk	0.575	0.604
Node2Vec	0.504	0.505
Random Forest	0.512	0.164
MultiGML-RGCN (basic)	1.0	1.0
MultiGML-RGAT (basic)	1.0	1.0
MultiGML-RGCN (multimodal)	1.0	1.0
MultiGML-RGAT (multimodal)	0.980	0.982

Table 3

Model performance results for predicting a novel gene - phenotype association. Test results of competing KG embedding methods, including TransE, RotatE, ComplEx, DistMult, DeepWalk and node2vec for prediction of a gene - phenotype association in comparison to our MultiGML models. Both MultiGML-RGCN and -RGAT variants were tested with basic and multimodal input features. The model variant “multimodal” refers to the use of several modalities for each node type described in section 3.1.2. In the model variant “basic” all input features have been initialized with the Xavier-Glorot method, i.e. the model effectively learns from the topology only.

Model	AUROC	AUPR
TransE	0.735	0.674
RotatE	0.723	0.680
ComplEx	0.843	0.770
DistMult	0.848	0.767
DeepWalk	0.654	0.630
Node2Vec	0.793	0.781
MultiGML-RGCN (basic)	0.898	0.832
MultiGML-RGAT (basic)	0.897	0.831
MultiGML-RGCN (multimodal)	0.897	0.832
MultiGML-RGAT (multimodal)	0.892	0.826

2.2. Use cases

To illustrate the practical use of our MultiGML method we further explored two newly predicted links between drugs and ADEs that were not part of the KG. Furthermore, we show an example of a newly predicted gene - phenotype association. All links have been predicted with probability >70% by MultiGML-RGAT.

2.2.1. Acute liver failure as a predicted adverse drug event of alendronic acid

MultiGML predicted a link between alendronic acid (DRUGBANK:DB00630), a bisphosphonate, and acute liver failure (UMLS: C0162557). Alendronic acid is used to prevent and treat osteoporosis [53], and was found to cause liver damage in a patient that was in treatment for osteoporosis [54]. Alendronic acid is also listed in the NIH LiverTox lexicon as a “rare cause of clinically apparent liver injury” [55].

To better explain the prediction by our model we investigated the attention coefficients calculated by the attention mechanism and the feature importances obtained via integrated gradients. First, we extracted the attention weights for all relations involving acute liver failure and alendronic acid of the MultiGML-RGAT model of the last graph attention layer (see Fig. 1 A). Several relations between alendronic acid and proteins, including two tyrosine phosphatases, PTPRS and PTPN4, and the phenotype Paget’s Disease (UMLS: C0029401), were weighted higher than all other direct relations by the MultiGML-RGAT model. Indeed, alendronic acid is used to treat Paget’s Disease of bone, also known as Osteitis Deformans [56] by inhibiting tyrosine-protein phosphatases [53]. Protein tyrosine phosphatase receptor type S (PTPRS) acts as a metastatic suppressor in hepatocellular carcinoma [57] and was found dysregulated in cirrhotic liver [58]. PTPN4 belongs to the same family of proteins and has accordingly been reported as a prognostic marker for hepatocellular carcinoma [59].

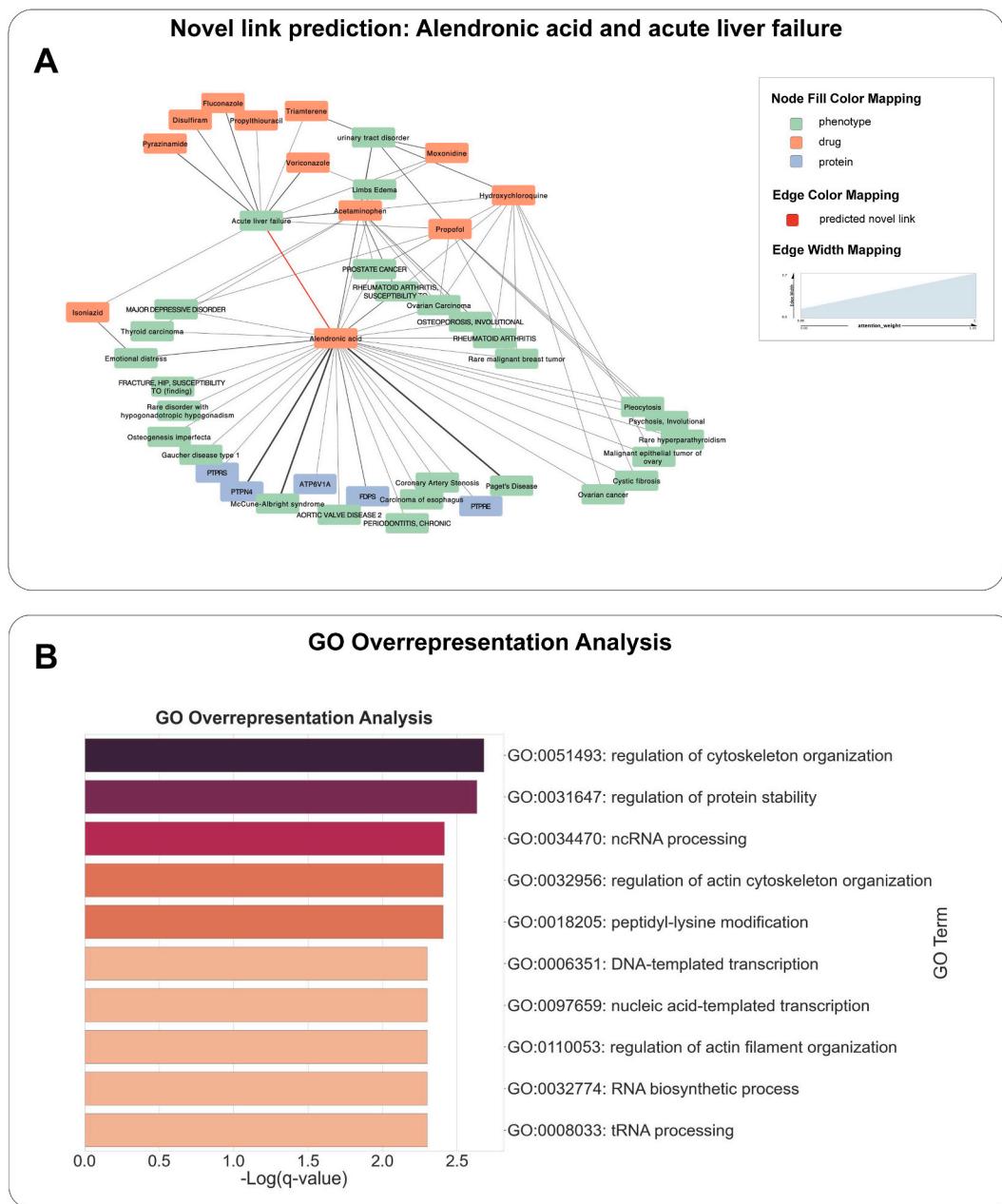


Fig. 1. Prediction of acute liver failure as an ADE for alendronic acid. A) Novel prediction of acute liver failure (UMLS:C0162557) as a potential ADE of alendronic acid (DRUGBANK:DB00630) in the KG, colored in red. The attention weight for every edge from the last MultiGML-RGAT graph attention layer is indicated by the edge strength. **B)** GO overrepresentation analysis of the top 100 genes from the L1000 drug signature identified via the integrated gradients method. Top 10 enriched terms (one per cluster) created with Metascape. $-\text{Log}_{10}(q)$ - values are reported and color coded for each term.

To better understand the prediction by our MultiGML model, we investigated the importances of the input features by using the integrated gradients method [60]. We focused on the gene expression signature of alendronic acid in our analysis. More specifically, we identified the top 100 influential genes of the L1000 gene expression signature of this drug. Next, we performed a Gene Ontology (GO) overrepresentation analysis via a hypergeometric test. After multiple testing correction, according to the Benjamini-Hochberg method and choosing a false discovery rate cutoff of 5% we identified the biological pathways that the most influential genes of the gene expression signature were enriched in. We found that regulation of the cytoskeleton organization and protein stability were important for the prediction of acute liver failure as an adverse drug event of alendronic acid (see Fig. 1 B, Suppl. Table 1). A comprehensive list of the top 100 positively and negatively attributed genes can be found in the Supplementary Material (Suppl. Table 2).

2.2.2. Paralysis as a predicted adverse drug event of kanamycin

Kanamycin (DRUGBANK:DB01172), which is an aminoglycoside bactericidal antibiotic [53], was predicted to be associated with a paralytic side effect (UMLS:C0522224). Kanamycin is used to treat a wide variety of bacterial infections [53]. In several studies, Kanamycin was reported to be neurotoxic and induce neuromuscular paralysis or blockades [61–63]. More recent studies with organoids suggested a damaging effect on early postnatal but not on adult ganglion neurons [64]. Indeed, ototoxicity of kanamycin is a significant dose-limiting side effect [65].

As done previously, we extracted the attention weights for all relations involving kanamycin and paralysis of the MultiGML-RGAT model of the last graph attention layer (see Fig. 2 A). Of all direct relations, the relation between the drug cyclopentolate and the phenotype paralysis was weighted higher than all other direct relations by the MultiGML-RGAT model. Cyclopentolate (DRUGBANK: DB00979) is an anticholinergic agent used to dilate the eye for diagnostic and examination purposes [53]. Additionally to inducing mydriasis - the dilation of the pupil -, cyclopentolate also causes reversible paralysis of the ciliary muscle by blocking muscarinic receptors [66].

We again investigated which genes from the gene expression signature of kanamycin were found to be important for the prediction by the integrated gradient analysis. We performed a GO overrepresentation analysis of the top 100 genes via a hypergeometric test (see Fig. 2 B, Suppl. Table 3), reporting the Benjamin-Hochberg adjusted *p*-value (*q*-value) for multiple testing. As a result, we found that biological processes involved in metabolism and responses to stimuli were significantly overrepresented in the most influential genes of the gene expression signature of kanamycin. Altogether, this demonstrates the ability of our method to point towards biological

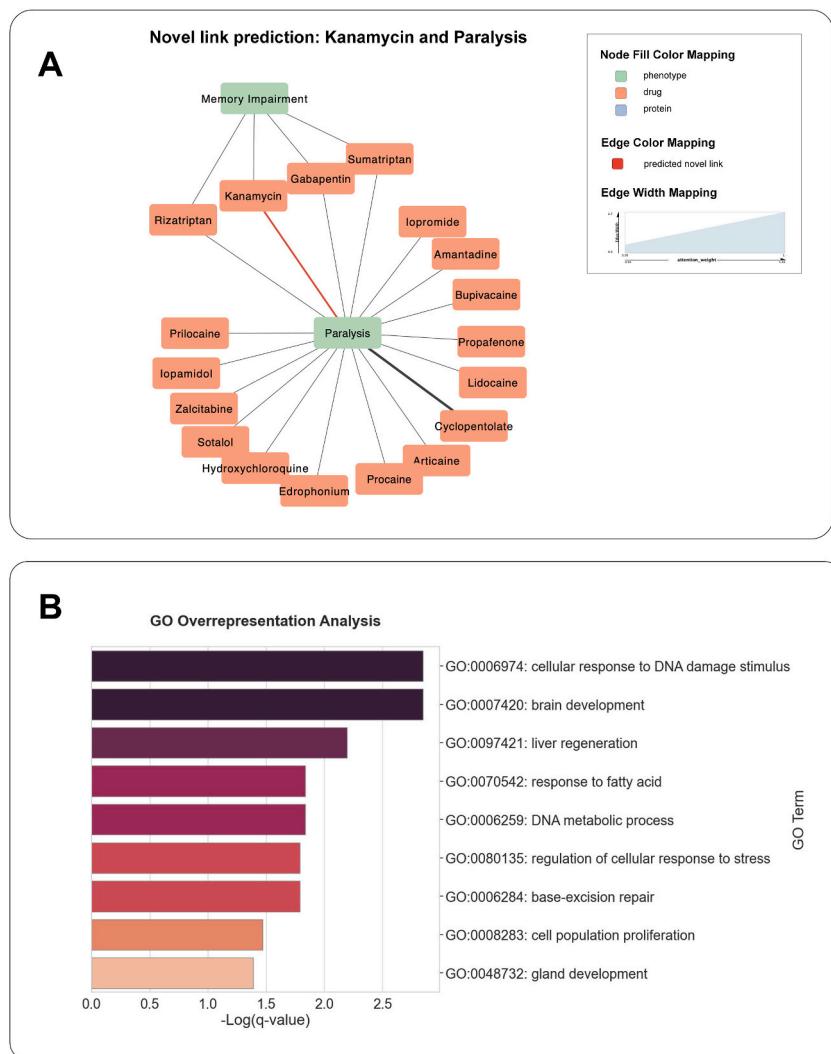


Fig. 2. Prediction of paralysis as an ADE for kanamycin. A) Novel prediction of paralysis (UMLS:C0522224) as potential ADE of kanamycin (DRUGBANK:DB01172) in the KG, colored in red. The attention weight for every edge from the last MultiGML-RGAT graph attention layer is indicated by the edge strength. B) GO overrepresentation analysis of the top 100 genes from the L1000 drug signature identified via the integrated gradients method. Top enriched terms created with Metascape. -Log₁₀(q) - values are reported and color coded for each term.

mechanisms associated with ADEs. A comprehensive list of the top 100 genes can be found in the Supplementary Material (Suppl. Table 4).

2.2.3. Association of WNT3 with thrombophlebitis

WNT3 (HGNC:12782) is a gene that is part of the Wnt signaling pathway (KEGG:hsa04310) in humans. Wnt proteins are secreted morphogens that are required for basic developmental processes, such as cell-fate specification, progenitor-cell proliferation and the control of asymmetric cell division, in many different species and organs [67]. Our MultiGML-RGAT model predicted an association between WNT3 and Thrombophlebitis (UMLS:C0040046). Thrombophlebitis is an inflammation of a vein associated with a blood clot [68]. Several studies suggest that WNT signaling has a regulatory role in inflammation [69], is involved in the calcification of vascular smooth muscle cells [70], and that it is a key player in the development of vascular disease, including thrombosis [71]. A recent study on endothelial injury has shown protective effects of Wnt signaling [72]. More specifically, attenuated apoptosis and exfoliation of vascular endothelial cells and infiltration of inflammatory cells was observed upon activation of the Wnt/beta-catenin pathway.

In Fig. 3, we display the novel link of WNT3 to thrombophlebitis in the knowledge graph together with the attention weights learned by our MultiGML-RGAT model. Several relations between WNT3 and other proteins were attributed a high attention weight, including insulin (INS), LRP6, as well as their relations with other proteins, FZD4, FZD7, FZD9, SFRP2 with INS and FZD10 with LRP6. Betamethasone and its relation to thrombophlebitis as an ADE was also attributed with a high attention weight. Associations between the phenotypes essential thrombocythemia (UMLS:C0040028) and rare diabetes mellitus (UMLS:C5681799) with insulin were also much attended by the model. The association between thrombosis, vascular inflammation and diabetes in relation with insulin resistance has been observed in many studies [73,74]. A modulation of the interaction of the insulin and Wnt signaling has even been proposed as an attractive target in treating diabetes [75], and could potentially have a role in mediating the effect of inflammatory conditions affecting the vascular system, such as thrombophlebitis. Despite this supporting evidence it should be highlighted that the exact link between WNT3 and thrombophlebitis is new and requires further clinical or experimental validation.

3. Materials and methods

3.1. Multi-modal knowledge graph generation

3.1.1. Integration of biomedical knowledge from databases

We integrated information from 14 well-established databases to generate an heterogeneous KG (see Table 4). Our KG contains information about interactions and associations between drugs, proteins, and phenotypes. We introduce the node type ‘phenotype’ to resolve the ambiguity between ADEs and diagnoses. That means ADEs and diagnoses are both subsumed as ‘phenotype’. As a result, we generated a heterogeneous and multi-relational KG with 3 different entity types and 8 different relation types (see Fig. 4A). It contains

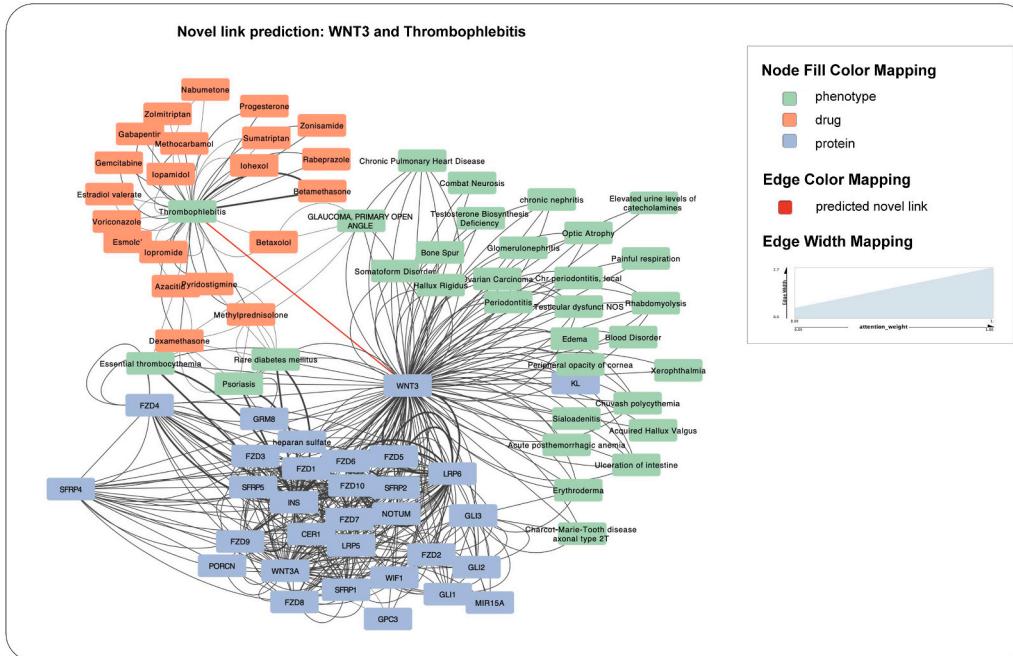


Fig. 3. Prediction of thrombophlebitis as a phenotype associated with WNT3. Novel prediction of thrombophlebitis (UMLS:C0040046) associated with WNT3 (HGNC:12782) in the KG, colored in red. The attention weight for every edge from the last MultiGML-RGAT graph attention layer is indicated by the edge strength.

Table 4

Source databases for knowledge graph generation. The databases that were used as a resource for building the heterogeneous and multi-relational knowledge graph in section 3.1. Are listed here with their total counts of relations that were selected.

Database	Count	Publication
BioGRID	102447	[76]
Clinical Trials	7626	[77]
DisGeNET	5448	[78]
DrugBank	10072	[53]
IntAct	23055	[79]
IUPHAR-DB	2379	[80]
KEGG	63356	[67]
NeuroMMSig	1761	[81]
OffSIDES	62	[82]
Open Targets	5222	[83]
Pathway Commons	32928	[84]
PheWAS Catalog	159202	[85]
Reactome	6379	[86]
SIDER	135	[87]

20,930 nodes and 420,072 relations (see Table 5 Table 6). The relation types that occur in the KG are drug-protein, protein-phenotype, genetic variant-phenotype, drug-adverse drug event and 3 different types of protein-protein interactions (physical interaction, functional interaction and signaling interaction). The knowledge graph is formally defined as $G = (V, L)$, with V as entities, and L as relations.

Interaction information between approved drugs and their protein targets was taken from DrugBank [53] and IUPHAR [80]. Associations between proteins and indications (here: phenotypes) with a high confidence score >0.6 were extracted from DisGeNet [78] and specific gene-phenotype associations from PheWAS [85], with an odds-ratio >1 . Drug indications for diseases were obtained from OpenTargets [83] and ClinicalTrials [77]. Protein-protein interactions were gathered from renowned databases, including KEGG [67,88], Reactome [86], BioGRID [76], IntAct [79], PathwayCommons [84] and NeuroMMSig [81]. The OFFSIDES database [82] as well as the renowned SIDER database [87] were used to extract known ADEs of drugs (see Table 4).

Because more severe ADEs tend to increase the risk for a drug to fail in clinical trials or be withdrawn from the market, the following heuristic was employed to filter out more severe ADEs from the information contained in the aforementioned databases: first, we designed a novel metric called the *failure ratio*, which we computed for each phenotype in the graph which served as a target node in an “adverse drug event” edge. The *failure ratio* is defined in Equation (1). Given the number of trials in which a given phenotype is listed as an “adverse drug event” on ClinicalTrials.gov, the *failure ratio* is equal to the proportion of these trials which were suspended, terminated, or withdrawn. All phenotypes with a failure ratio greater than 0.75 among at least 3 trials were chosen to be used for the graph. This resulted in nearly two hundred ADE relations, approximately 0.1% of the unique ADE between SIDER and OFFSIDES, which were subsequently added to the KG. All identifiers and interaction types were harmonized across all databases (Fig. 4 A). A complete overview about the entity and relation types in the KG is provided in Tables 5 and 6.

$$\text{Failure ratio} = (\text{Trials with condition as "adverse event" AND suspended, terminated or withdrawn}) / (\text{Total trials with condition as "adverse event"}) \quad (1)$$

3.1.2. Definition of entity related features

The integration of multiple biologically, chemically and medically relevant modalities into a knowledge graph enriches the information quality of the graph. This enrichment may subsequently be beneficial for downstream link prediction tasks as well the post-hoc explanation of neural network models. Therefore, we decided to incorporate multiple feature modalities in our dataset (see Fig. 4 B). We chose modalities that were descriptive for the individual entity type, and generated features for each entity type as described below:

- **DRUGS:** Transcriptomics data are informative about the effect of a drug on biological processes in a defined system of a cell culture experiment. A molecular signature can therefore be generated for each drug, measuring the gene expression fold change of selected transcripts. We chose the LINCS L1000 dataset [89] to annotate the drugs with gene expression profile information. More specifically, we retrieved the consensus signatures calculated by Himmelstein et al. [21,90]. The background is that each LINCS compound may have been assayed across multiple cell lines, dosages and replicates. Himmelstein et al. thus estimated a single consensus transcriptional profile across multiple signatures.

The effects of a drug perturbation in a cell culture experiment can not only be seen in the gene expression fold change, but also in the change in morphology of the treated cells. Therefore, we additionally annotate the drug with the Cell Painting morphological profiling assay information from the LINCS Data Portal (LDG-1192: LDS-1195) [91].

Furthermore, the molecular structure of the drugs was also taken into account by generating the molecular fingerprints. We here took the Morgan count fingerprint [92] with a radius = 2, generated with the RDKit [93].

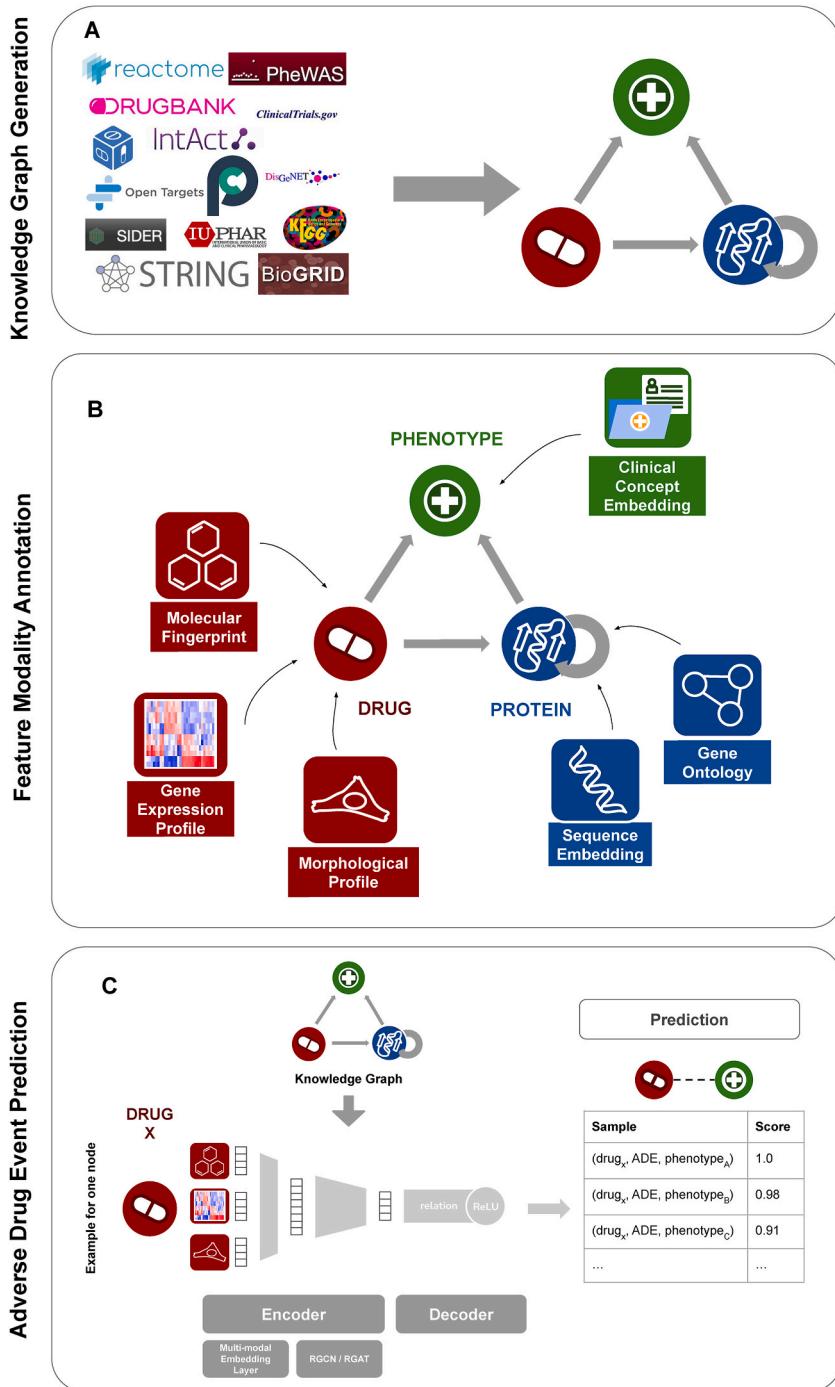


Fig. 4. Overview of workflow. A) Knowledge Graph compilation. In the first step of data processing, interaction information from 14 biomedical databases was parsed with data on drug-drug interactions, drug-target interactions, protein-protein interactions, indication, drug-ADE and gene-phenotype associations. The data was harmonized across all databases and a comprehensive, heterogeneous, multi-relational knowledge graph was generated. **B) Feature definition.** Descriptive data modalities were selected to annotate entities in the knowledge graph. Drugs were annotated with their molecular fingerprint, the gene expression profile they cause, and the morphological profile of cells they induce. Proteins were annotated with their protein sequence embedding and a gene ontology fingerprint. Phenotypes, comprising indications and ADEs, were annotated by their clinical concept embedding. **C) Proposed MultiGML approach.** The heterogeneous Knowledge Graph with its feature annotations is used as the input for our graph neural network approach, the MultiGML. For each node entity, a multi-modal embedding layer learns a low dimensional representation of entity features. These embeddings are then used as input for either the RGCN or RGAT of the encoder (see section 3.2.1), which learns an embedding for each entity in the KG. A bilinear decoder takes a source and a destination node, drug X and several phenotypes A, B and C in the example here, and produces a score for the probability of their connection, considering their relation type with each other.

Table 5
Overview of entities in the knowledge graph.

Entity Type	Count
phenotype	16,560
drug	2378
protein	12,953

Table 6
Overview of relation types in the knowledge graph.

Relation Type	Count
drug-adverse drug event	197
functional protein-protein association	1761
protein-phenotype	5448
physical protein-protein interaction	6087
drug-protein	12,451
drug-indication	12,848
genetic variant-phenotype	159,202
protein-protein signaling interaction	222,078

- **PROTEINS:** We used structural information of proteins in form of protein sequence embeddings. We generated the embeddings for each protein with the ESM-1b Transformer [94], a recently published pre-trained deep learning model for protein sequences.

In addition, we generated a binary Gene Ontology (GO) fingerprint for biological processes for each protein using data from the Gene Ontology Resource [95,96]. A total of 12,226 human GO terms of Biological Processes were retrieved and their respective parent terms obtained. This resulted in a 1298 dimensional binary fingerprint for each protein, with each index either set to 1, if the protein was annotated with the respective GO term or 0 if not.

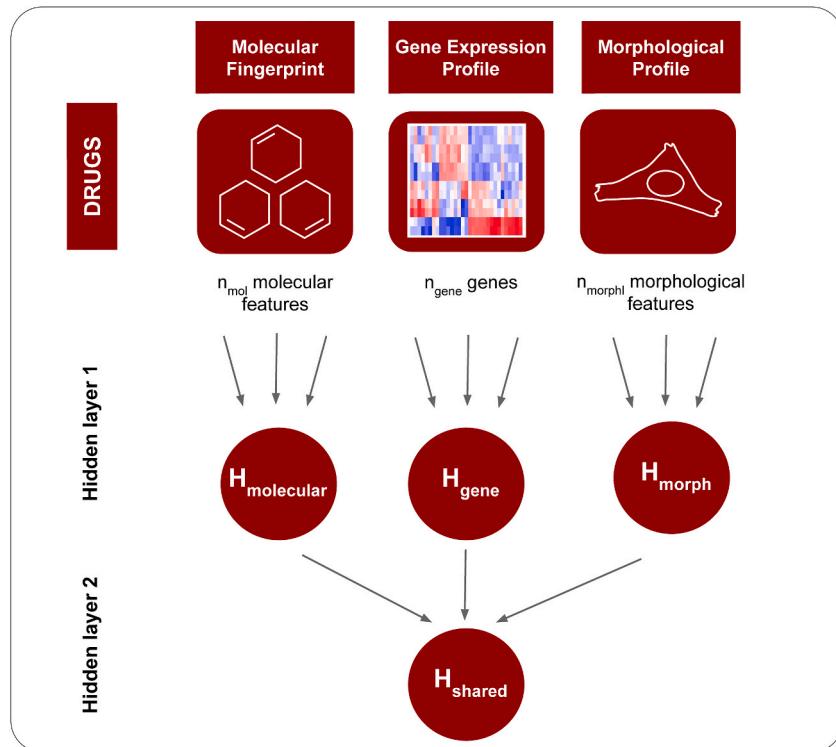


Fig. 5. Multi-modal embedding (example of drug input): Each drug is represented by f different feature modalities, which are fed into a multi-modal neural network with bottleneck architecture. That means $H_{molecular}$, H_{gene} , H_{morph} are the output of dense feed-forward layers, each having $k_r / 2$ hidden units, where k_r is the number of original input features for data modality r . $H_{shared} = (H_{molecular} || H_{gene} || H_{morph})$ represents the multi-modal embedding.

- **PHENOTYPES:** Medical concept embeddings from Beam et al. [97] were used to annotate phenotypes including ADEs and indications. The so-called *cui2vec* embeddings were generated on the basis of clinical notes, insurance claims, and biomedical full text articles for each clinical concept. Briefly, the authors mapped ICD-9 codes in claims data to UMLS concepts and then counted co-occurrence of concept pairs. After decomposing the co-occurrence matrix via singular value decomposition, they used the popular word2vec approach [98] to obtain concept embeddings in the Euclidean space. We refer to Beam et al. for more details.

3.2. Graph neural network architecture

MultiGML consists of two main structures, an encoder and a decoder. The encoder has two main components which create a low-dimensional embedding of each node in the KG (see section 3.2.1.). Due to its design, the encoder can handle multimodal input data. The second part of the model decodes edges from node embeddings with a bilinear form (see section 3.2.2.). The entire model architecture is shown in Fig. 4C and discussed in more detail in the subsequent paragraphs.

3.2.1. Encoder

3.2.1.1. Multi-modal embedding of node features. Due to the multi-modal character of our KG, we require a model that can integrate input features of multiple modalities for one node into the message passing. To do so, we implemented a specific architecture based on our previous work [99] that combines representations from different data modalities (Fig. 5).

In a nutshell, this multimodal embedding learns hidden representations of each modality separately in the first densely connected layer. The hidden feature representations are then concatenated and passed to a second densely connected layer to generate a shared multimodal embedding for each entity v in the KG: Let x_1, x_2, \dots, x_k denote the k feature vectors of dimensions d_1, d_2, \dots, d_k associated to entity v . The embedding H_{shared} (see Equation (2)) is therefore learned as follows:

$$H_{\text{shared}} = \sigma(W_{\text{multi}}(\sigma(W_1(x_1)) \parallel \sigma(W_2(x_2)) \parallel \dots \parallel \sigma(W_k(x_k)))) \quad (2)$$

where σ is the *tanh* activation function and \parallel denotes a concatenation.

We use dropout units in each layer with a dropout ratio that is adjusted during Bayesian hyperparameter optimization. This is followed by a batch normalization with a *tanh* activation function.

3.2.1.2. Relational Graph Convolutional Neural Network (RGCN). KGs often incorporate a variety of heterogeneous information in the form of different node and edge types. In the following, we will refer to the prominent Relational Graph Convolutional Neural Network (RGCN) that was proposed by Schlichtkrull et al. [100] to deal with the multirelational data characteristic of KGs. The RGCN includes information from the neighborhood of a node into the message passing by differentiating between the relation types. Due to this characteristic, the model is able to learn the inherent relationships between the entities in the KG.

The RGCN takes as input a heterogeneous multi-relational knowledge graph G with features $x \in \mathbb{R}^q$ and learns an embedding h_i of each entity $v_i \in V$ in the KG. The architecture of the implemented model has three consequent RGCN layers¹⁰⁰:

- **input to hidden layer:** input feature vectors $x_i \in \mathbb{R}^q$ are transformed into their hidden representation $h_i \in \mathbb{R}^{k'}$
- **hidden to hidden layer:** convolution of hidden feature vectors $h_i \in \mathbb{R}^{k'}$, maintaining their shape
- **hidden to output layer:** hidden feature vectors $h_i \in \mathbb{R}^{k'}$ are transformed into their latent representation $h_i \in \mathbb{R}^1$

The message passing for node i is given by Equation (3):

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)} \right) \quad (3)$$

The updated hidden representation h_i of entity v_i at layer $l+1$ is a non-linear combination of the hidden representations of neighboring entities with index $j \in N_i^r$ weighted by the learnable relation type specific weight matrix $W_r^{(l)}$. Here, N_r is the set of neighbors of node v_i of relation type r . A self-loop is defined by adding the node's own hidden representation h_i , multiplied by the weight matrix W_0 . $c_{i,r}$ is a normalization constant that is task-dependent and can either be learned or chosen in advance, such as $c_{i,r} = |N_i^r|$. We refer to this variant of our MultiGML model as MultiGML-RGCN.

3.2.1.3. Relational graph attention network (RGAT). Alternatively to the RGCN we considered a relational graph attention network [42] as part of the encoder. The input is a set of entity features $h = \{h_1, h_2, h_3, \dots, h_V\}, h_i \in \mathbb{R}^p$ with V being the number of entities and p being the number of features of each entity. Self-attention is performed on the entities, whereby a shared attention mechanism a computes attention coefficients [42] for each relation type (Equation (4))

$$E_{i,j}^{(r)} = a(W^{(r)} h_i, W^{(r)} h_j) \quad (4)$$

The attention mechanism a is a single-layer feedforward neural network feeding into a Leaky ReLU unit (angle of negative slope = 0.2). The attention coefficients are normalized across all choices of j via the softmax function [42,101] (Equation (5)),

$$\alpha_{i,j}^{(r)} = \text{softmax}_j\left(E_{i,j}^{(r)}\right) = \frac{\exp\left(E_{i,j}^{(r)}\right)}{\sum_{k \in N_i^{(r)}} \exp\left(E_{i,k}^{(r)}\right)}, \forall i, r : \sum_{j \in N_i^{(r)}} \alpha_{i,j}^{(r)} = 1 \quad (5)$$

leading to the propagation model for a single node update in a multi-relational graph of the following form (Equation (6)):

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^{(r)}} \alpha_{i,j}^{(r)} h_j \left(W^{(r)} \right)^T \right) \quad (6)$$

The graph attention layer allows assigning different importances to nodes of a same neighborhood which can be analyzed to interpret the model predictions [42]. We refer to this variant of our MultiGML model as MultiGML-RGAT.

3.2.2. Bilinear decoder

The decoder structure in our model uses the entity embeddings to decode relations in the KG. We calculate a score for a given triple of entities v_i and v_j connected by relation r . We use a bilinear form on the embeddings h_i and h_j with a trainable matrix M_r representing the relation type and apply a sigmoid function σ to the result as in Equation (7):

$$\text{score}_{i,j}^{(r)} = \text{score}(v_i, r, v_j) = \sigma(h_i^T M_r h_j) \quad (7)$$

3.3. Empirical evaluation

3.3.1. Model training strategy

We trained our MultiGML model on the KG described in section 3.1. The binary cross-entropy loss was applied to supervise the model. We performed a stratified random split of all relations into a 70% train, 20% test and 10% validation set. The stratification was done such that each data subset contained the same fraction of each relation type. The number of relations amounted to 302,445 in the training set, 33,608 in the validation set and 84,019 in the test set. Note that all those real existing relations provide positive examples. We wanted to see whether the input features have an effect on the performance of the models, and therefore ran our experiments with different types of input features. We created random uniform features that do not express any biological meaning, which we refer to as the “basic” feature variant, and we applied the multi-modal biological, chemical and medical features described in section 3.1.2. to the “multi-modal” feature variant. An important question in Graph Machine Learning is, how to generate negative samples for non-existing relations. In this work we performed negative sampling where for each positive relation one negative relation was generated by randomly exchanging the target for each source entity according to a uniform distribution (uniform sampling). In very small graphs it can happen due to this sampling technique that randomly generated negative samples actually represent a true positive sample. In large graphs such as ours, this is however not of concern since this would be a rare event. Yet, we made sure when evaluating the predictions that no true positive sample was counted as a negative sample. We used the Deep Graph Library (DGL) Python package [102] to implement the graph neural networks, the PyTorch package [103] for the multi-modal embedding layer, and applied the PyTorch Lightning framework [104] for high-performance artificial intelligence. We employed a hyperparameter optimization with the Optuna package [105], with a customized search space for each hyperparameter (see Suppl. Table 5). Notably, hyperparameter optimization also included the selection of node feature modalities. That means we allowed entire data modalities to be dropped from the model during training. The Tree Parzen Estimator [106], an independent sampling algorithm, was used for an efficient exploration of the search space. Each hyperparameter optimization for both the RGCN and RGAT model consisted of 50 trials. Models within each trial were trained for 100 epochs unless the hyperband pruner [107] determined that a trial should be pruned. After each training epoch the model was evaluated on the validation test. After the best set of hyperparameters was found, we trained a final model for 100 epochs with the selected hyperparameters (see Suppl. Table 7). The problem of overfitting (i.e. low bias and high variance of model predictions) was counteracted by several strategies. First of all, a large training set of more than 300.000 samples was used, and we ensured a stratified split of relation types into train, validation and test set to provide a dataset with high variability and low bias. During training of the model, L2 regularization was used. Furthermore, early stopping was employed, triggered in case of a stagnation in the loss calculation for 10 epochs. We also tried to reduce the complexity of the model architecture by limiting the number of hidden layers to max. 7 during the hyperparameter optimization. Finally, the model was tested on the unseen test set.

3.3.2. Comparison against competing methods

We benchmarked MultiGML against several competing approaches for both general link prediction and adverse drug event prediction. All methods were evaluated using the same data splits. We compared our models against four well-established KG embedding approaches, namely TransE [34], RotatE [37], DistMult [36] and ComplEx [35], DeepWalk [38] and node2vec [39]. All of these models produce an embedding of a KG. Shortly summarized, TransE models relations as translations of a source entity v_{source} , a target entity v_{target} and a relation r in the embedding space by trying to minimize the distance $d(v_{source} + r, v_{target})$, while RotatE represents relations as a rotation from the source entity to the target entity - both are translation-based approaches. DistMult and ComplEx both use similarity-based scoring functions, where DistMult restricts r to be a diagonal matrix $diag$, and ComplEx extends DistMult to the complex vector space. DeepWalk and node2vec are KG embedding techniques based on random walks. DeepWalk creates a mapping of entities to a low-dimensional space of features using a random walk strategy, where neighboring nodes have equal probability to be

Table 7

Summary of competing methods that were used as a comparison to our MultiGML variants.

Model	Type of model	Reference
TransE	translation-based	[34]
RotatE	translation-based	[37]
DistMult	semantic matching-based	[36]
ComplEx	semantic matching-based	[35]
DeepWalk	random walk-based	[38]
Node2vec	random walk-based	[39]
Random Forest	ensemble-based	[10]

chosen at the next step, while node2vec additionally uses weights that influence the random walk behavior. The entity embeddings and the scores for all samples were generated using the implementation by Ref. [108], which incorporates a multi-layer perceptron with three layers as a predictor. We also compared our models to a Random Forest (RF) based machine learning approach for ADE prediction [10], which uses gene expression as data as well as compound fingerprints. The competing methods are summarized in Table 7.

3.4. Making models explainable

From an application perspective it is important to be able to explain which features of a compound influence the model's predictions in each single instance. For this purpose we build on the integrated gradients method [60] as implemented in captum.ai [109]. Integrated gradients is an axiomatic attribution method which represents the integral of gradients with respect to inputs along the path from a given baseline to input [60]. The integrated gradient along the i th dimension from baseline x_0 to input x is defined in Equation (8):

$$\text{IntegratedGrads}_i(x)^{\text{approx}} := (x_i - x'_i) \times \int_0^1 \frac{\delta F(x' + \alpha \times (x - x'))}{\delta x_i} d\alpha \quad (8)$$

with α as a scaling coefficient and F as a function $F : R^p \Rightarrow (0, 1)$ which represents our MultiGML model. The Integrated Gradients method provides information about both local and global feature contributions. Local feature contributions can be explained by the completeness axiom, which states that "given x and a baseline x_0 , the attributions of x add up to the difference between the output of F at the input x and the baseline x_0 " is chosen. In case both are true, all attributions are on the same scale and can be compared globally with each other.

For each predicted link between a drug and a side effect, we calculated the integrated gradients to receive the importances of the individual features. We used $n = 500$ steps for the approximation method. We focused our analysis on the gene expression signature of the drug entities. A mean vector was used for the baseline x_0 . In a further step, the attributions of the Integrated Gradient analysis were evaluated. The top 100 influential genes of the gene expression signature were identified for each drug in the ADE prediction. As a next step, a gene ontology enrichment analysis of biological processes was performed on the top 100 positively and negatively attributed genes from the molecular gene expression signature of the drugs using Metascape [110]. For the enrichment analysis, all genes in the human genome were used as background, and a q -value cut-off of 0.05, a minimum count of 3 and an enrichment factor >1.5 were chosen. A hypergeometric test was performed and q -values were calculated using the Benjamini-Hochberg procedure [111] to account for multiple testing.

We evaluate our models on the independent test set according to area under the ROC curve (AUROC) and area under the precision-recall curve (AUPR).

4. Conclusions

We proposed a novel Graph Machine Learning neural network architecture for adverse drug event prediction that combines multi-modal quantitative data with a heterogeneous, multi-relational KG. MultiGML uses a multi-modal encoder to learn an embedding of multiple input data modalities into a joint space. Each point in this joint space represents a node of the graph. Subsequently, we use heterogeneous graph convolution and graph attention techniques, respectively, to consider the knowledge graph structure. Finally, a bilinear decoder is employed for link prediction.

MultiGML demonstrated excellent prediction performance in comparison to a broad set of competing approaches, including translation based (TransE, RotatE), semantic matching based (DistMult, ComplEx) and random walk based (DeepWalk, node2vec) techniques. Furthermore, we demonstrated that predictions made by our MultiGML method could be explained via the method of integrated gradients and visualization of attention weights. We showed that due to the integration of multimodal node features it was possible to identify biologically plausible mechanisms associated with predicted ADEs. Therefore, our approach could provide valuable information during the early phases of drug development, where it is important to lower the failure risk of later clinical trials. Getting insights into relevant biological mechanisms associated with a high risk could support selection of safe targets. We thus see the value of integrating multimodal data into MultiGML not so much in terms of increase in link prediction performance, but much more with regard to the far better interpretability compared to purely graph topology based techniques.

5. Limitations

Of course, MultiGML is not without limitations: for example, specific feature modalities may be unavailable for some of the entities in the KG. The neighborhood aggregation approach of MultiGML in such a case provides a way to mitigate this issue, because it essentially smoothens features over the neighborhood of a given entity, but that can not perfectly replace missing information. Furthermore, ADEs are in reality also dependent on pharmacodynamic (PD) properties of a compound, including dose, which are currently not considered in our model. Finally, model explanations can only disentangle model predictions, but they do not always point to the right biological cause of an ADE. Due to the versatile model design, MultiGML offers the perspective of applications of link prediction tasks other than the one discussed in this paper, including drug repositioning. Moreover, MultiGML could be used to integrate other or additional data modalities, for example protein and tissue expression and pathology imaging slides. Altogether, we thus see MultiGML as a flexible approach to support important decisions in early drug discovery.

Author contribution statement

Holger Fröhlich, Andrea Zaliani: Conceived and designed the analysis; Analyzed and interpreted the data; Wrote the paper.
 Sheraz Gul: Analyzed and interpreted the data; Wrote the paper.
 Sophia Krix, Lauren Nicole DeLong, Daniel Domingo-Fernandez, Sumit Madan, Ashar Ahmad: Analyzed and interpreted the data; Contributed analysis tools or data; Wrote the paper.

Data availability statement

Data associated with this study has been deposited at <https://github.com/SCAI-BIO/MultiGML>.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: DDF received salaries from Enveda Biosciences, and AA from Grünenthal GmbH. Both companies had no influence on the scientific results reported in this paper.

Acknowledgements

We thank Bruce Schultz and Aliaksandr Masny for their support during the project. This work was supported by the Research Center Machine Learning (FZML) of the Fraunhofer Cluster of Excellence Cognitive Internet Technologies CCIT.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e19441>.

References

- [1] J.R. Nebeker, P. Barach, M.H. Samore, Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting, *Ann. Intern. Med.* 140 (2004) 795–801.
- [2] K.J. Carss, et al., Using human genetics to improve safety assessment of therapeutics, *Nat. Rev. Drug Discov.* (2022) 1–18, <https://doi.org/10.1038/s41573-022-00561-w>.
- [3] A. Duffy, et al., Tissue-specific genetic features inform prediction of drug side effects in clinical trials, *Sci. Adv.* 6 (2020), eabb6242.
- [4] P.A. Nguyen, D.A. Born, A.M. Deaton, P. Nioi, L.D. Ward, Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects, *Nat. Commun.* 10 (2019) 1579.
- [5] M. Liu, et al., Determining molecular predictors of adverse drug reactions with causality analysis based on structure learning, *J. Am. Med. Inf. Assoc.* 21 (2014) 245–251.
- [6] Y. Niu, W. Zhang, Quantitative prediction of drug side effects based on drug-related features, *Interdiscip Sci* 9 (2017) 434–444.
- [7] E. Pauwels, V. Stoven, Y. Yamanishi, Predicting drug side-effect profiles: a chemical fragment-based approach, *BMC Bioinf.* 12 (2011) 169.
- [8] Y. Yamanishi, E. Pauwels, M. Kotera, Drug side-effect prediction based on the integration of chemical and biological spaces, *J. Chem. Inf. Model.* 52 (2012) 3284–3292.
- [9] W. Zhang, F. Liu, L. Luo, J. Zhang, Predicting drug side effects by multi-label learning and ensemble learning, *BMC Bioinf.* 16 (2015) 365.
- [10] X. Zhao, L. Chen, J. Lu, A similarity-based method for prediction of drug side effects with heterogeneous information, *Math. Biosci.* 306 (2018) 136–144.
- [11] S.N. Deftereos, C. Andronis, E.J. Friedla, A. Persidis, A. Persidis, Drug repurposing and adverse event prediction using high-throughput literature analysis, *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3 (2011) 323–334.
- [12] A. Cakir, M. Tuncer, H. Taymaz-Nikerel, O. Ulucan, Side effect prediction based on drug-induced gene expression profiles and random forest with iterative feature selection, *Pharmacogenomics J.* 21 (2021) 673–681.
- [13] Z. Wang, N.R. Clark, A. Ma'ayan, Drug-induced adverse events prediction with the LINCS L1000 data, *Bioinformatics* 32 (2016) 2338–2345.
- [14] M. Takarabe, M. Kotera, Y. Nishimura, S. Goto, Y. Yamanishi, Drug target prediction using adverse event report systems: a pharmacogenomic approach, *Bioinformatics* 28 (2012) i611–i618.
- [15] S. Vilar, R. Harpaz, L. Santana, E. Uriarte, C. Friedman, Enhancing adverse drug event detection in electronic health records using molecular structure similarity: application to pancreatitis, *PLoS One* 7 (2012), e41471.

- [16] P. Schotland, et al., Target adverse event profiles for predictive safety in the postmarket setting, *Clin. Pharmacol. Ther.* 109 (2021) 1232–1243.
- [17] A.-L. Barabási, Z.N. Oltvai, Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.* 5 (2004) 101–113.
- [18] T. Rebele, et al., YAGO: a multilingual knowledge base from wikipedia, wordnet, and geonames, in: P. Groth, et al. (Eds.), *The Semantic Web – ISWC 2016*, Springer International Publishing, 2016, 9982 177–185.
- [19] D. Vrandečić, M. Wikidata Krötzsch, A free collaborative knowledgebase, *Commun. ACM* 57 (2014) 78–85.
- [20] A. Breit, S. Ott, A. Agibetov, M. Samwald, OpenBioLink: a benchmarking framework for large-scale biomedical link prediction, *Bioinformatics* 36 (2020) 4097–4098.
- [21] D.S. Himmelstein, et al., Systematic integration of biomedical knowledge prioritizes drugs for repurposing, *Elife* 6 (2017).
- [22] S. Zheng, et al., PharmKG: a Dedicated Knowledge Graph Benchmark for Biomedical Data Mining, *Brief Bioinform.* 2020.
- [23] Z. Chen, et al., CTKG: A Knowledge Graph for Clinical Trials, 2021, <https://doi.org/10.1101/2021.11.04.21265952>, 2021.11.04.21265952 Preprint at.
- [24] J. Lin, et al., Prediction of adverse drug reactions by a network based external link prediction method, *Anal. Methods* 5 (2013) 6120–6127.
- [25] Y. Luo, Q. Liu, W. Wu, F. Li, X. Bo, Predicting drug side effects based on link prediction in bipartite network, in: 2014 7th International Conference on Biomedical Engineering and Informatics, 2014, pp. 729–733, <https://doi.org/10.1109/BMEI.2014.7002869>.
- [26] D.M. Bean, et al., Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records, *Sci. Rep.* 7 (2017), 16416.
- [27] A. Cami, A. Arnold, S. Manzi, B. Reis, Predicting adverse drug events using pharmacological network models, *Sci. Transl. Med.* 3 (2011), 114ra127.
- [28] J.K. Aronson, *Meyley's Side Effects of Drugs: the International Encyclopedia of Adverse Drug Reactions and Interactions*, Elsevier, 2015.
- [29] A. Fokoue, M. Sadoghi, O. Hassanzadeh, P. Zhang, Predicting drug-drug interactions through large-scale similarity-based link prediction, in: H. Sack, et al. (Eds.), *The Semantic Web. Latest Advances and New Domains*, Springer International Publishing, 2016, pp. 774–789, https://doi.org/10.1007/978-3-319-34129-3_47.
- [30] Md R. Karim, et al., Drug-drug interaction prediction based on knowledge graph embeddings and convolutional-LSTM network, in: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics 113–123, Association for Computing Machinery, 2019, <https://doi.org/10.1145/3307339.3342161>.
- [31] P. Joshi, M. V. A. Mukherjee, A knowledge graph embedding based approach to predict the adverse drug reactions using a deep neural network, *J. Biomed. Inf.* 132 (2022), 104122.
- [32] W. Zhang, et al., Feature-derived graph regularized matrix factorization for predicting drug side effects, *Neurocomputing* 287 (2018) 154–162.
- [33] W. Zhang, et al., Predicting potential side effects of drugs by recommender methods and ensemble learning, *Neurocomputing* 173 (2016) 979–987.
- [34] 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).
- [35] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, G. Bouchard, Complex Embeddings for Simple Link Prediction, 2016, <https://doi.org/10.48550/arxiv.1606.06357> arXiv, https://papers.nips.cc/paper_files/paper/2013/hash/1ecc7a77928ca8133fa24680a88d2f9-Abstract.html.
- [36] B. Yang, W. Yih, X. He, J. Gao, L. Deng, Embedding Entities and Relations for Learning and Inference in Knowledge Bases, 2014 arXiv e-prints, <https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6575Y>.
- [37] Z. Sun, Z.-H. Deng, J.-Y. Nie, J. Tang, RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space, 2019, <https://doi.org/10.48550/arxiv.1902.10197> arXiv.
- [38] B. Perozzi, R. Al-Rfou, S. Skiena, DeepWalk: online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '14 701–710, ACM Press, 2014, <https://doi.org/10.1145/262330.2623732>.
- [39] A. Grover, J. Leskovec, node2vec, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864, <https://doi.org/10.1145/2939672.2939754>.
- [40] F. Scarselli, M. Goril, A.C. Tsai, M. Hagenbuchner, G. Monfardini, The graph neural network model, *IEEE Trans. Neural Network.* 20 (2009) 61–80.
- [41] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, Preprint at, <https://doi.org/10.48550/arxiv.1609.02907>, 2017.
- [42] P. Veličković, et al., Graph Attention Networks, 2017, <https://doi.org/10.48550/arxiv.1710.10903> arXiv.
- [43] C. Wang, S. Pan, G. Long, X. Zhu, J. Jiang, MGAE: marginalized graph autoencoder for graph clustering, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management - CIKM '17 889–898, ACM Press, 2017, <https://doi.org/10.1145/3132847.3132967>.
- [44] S. Rhee, S. Seo, S. Kim, Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification, in: J. Lang (Ed.), *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, 2018, <https://doi.org/10.24963/ijcai.2018/490>, 3527–3534.
- [45] D. Duvenaud, et al., Convolutional Networks on Graphs for Learning Molecular Fingerprints, 2015, <https://doi.org/10.48550/arxiv.1509.09292> arXiv.
- [46] A.M. Fout, *Protein Interface Prediction Using Graph Convolutional Networks*, Colorado State University, Libraries, 2016.
- [47] Y. Wu, M. Gao, M. Zeng, J. Zhang, M. BridgeDPI. Li, A novel graph neural network for predicting drug-protein interactions, *Bioinformatics* (2022), <https://doi.org/10.1093/bioinformatics/btac155>.
- [48] M. Zitnik, M. Agrawal, J. Leskovec, Modeling polypharmacy side effects with graph convolutional networks, *Bioinformatics* 34 (2018) i457–i466.
- [49] H. Kwak, et al., Drug-disease graph: predicting adverse drug reaction signals via graph neural network with clinical data, Preprint at, <http://arxiv.org/abs/2004.00407>, 2020.
- [50] L. Yu, M. Cheng, W. Qiu, X. Xiao, W. Lin, idse-HE: hybrid embedding graph neural network for drug side effects prediction, *J. Biomed. Inf.* 131 (2022), 104098.
- [51] P. Xuan, et al., Integrating specific and common topologies of heterogeneous graphs and pairwise attributes for drug-related side effect prediction, *Briefings Bioinf.* 23 (2022), bbac126.
- [52] D. Schuster, C. Laggner, T. Langer, Why drugs fail—a study on side effects in new chemical entities, *Curr. Pharmaceut. Des.* 11 (2005) 3545–3559.
- [53] D.S. Wishart, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res.* 46 (2018) D1074–D1082.
- [54] A. Halabe, B.M. Lifschitz, J. Azuri, Liver damage due to alendronate, *N. Engl. J. Med.* 343 (2000) 365–366.
- [55] J.H. Hoofnagle, J. Serrano, J.E. Knoben, V.J. Navarro, LiverTox: a website on drug-induced liver injury, *Hepatology* 57 (2013) 873–874.
- [56] I.R. Reid, E. Siris, Alendronate in the treatment of Paget's disease of bone, *Int. J. Clin. Pract. Suppl.* 101 (1999) 62–66.
- [57] Z.-C. Wang, et al., Protein tyrosine phosphatase receptor S acts as a metastatic suppressor in hepatocellular carcinoma by control of epidermal growth factor receptor-induced epithelial-mesenchymal transition, *Hepatology* 62 (2015) 1201–1214.
- [58] K.-M. Chan, et al., Bioinformatics microarray analysis and identification of gene expression profiles associated with cirrhotic liver, *Kaohsiung J. Med. Sci.* 32 (2016) 165–176.
- [59] G. Zhangyuan, et al., Prognostic value of phosphotyrosine phosphatases in hepatocellular carcinoma, *Cell. Physiol. Biochem.* 46 (2018) 2335–2346.
- [60] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks, Preprint at, <http://arxiv.org/abs/1703.01365>, 2017.
- [61] F.R. Freeman, Unusual neurotoxicity of kanamycin, *JAMA* 200 (1967) 410.
- [62] J.G. Naiman, K. Sakurai, J.D. Martin, The antagonism of calcium and neostigmine to kanamycin-induced neuromuscular paralysis, *J. Surg. Res.* 5 (1965) 323–328.
- [63] C.B. Pittenger, Y. Eryasa, R. Adamson, Antibiotic-induced paralysis, *Anesth. Analg.* 49 (1970) 487–501.
- [64] K. Gao, D. Ding, H. Sun, J. Roth, R. Salvi, Kanamycin damages early postnatal, but not adult spiral ganglion neurons, *Neurotox. Res.* 32 (2017) 603–613.
- [65] S.K. Heysell, et al., Hearing loss with kanamycin treatment for multidrug-resistant tuberculosis in Bangladesh, *Eur. Respir. J.* 51 (2018).
- [66] *Clinical Ocular Pharmacology*, Butterworth-Heinemann/Elsevier, 2008.
- [67] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* 44 (2016) D457–D462.
- [68] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, *Nucleic Acids Res.* 32 (2004) D267–D270.
- [69] S.J. George, Wnt pathway, *Arterioscler. Thromb. Vasc. Biol.* 28 (2008) 400–402.
- [70] K. Bundy, J. Boone, C.L. Simpson, Wnt signaling in vascular calcification, *Front. Cardiovasc. Med.* 8 (2021).

- [71] S. Foulquier, et al., WNT signaling in cardiac and vascular disease, *Pharmacol. Rev.* 70 (2018) 68–141.
- [72] Y. Wang, et al., Study on protection of human umbilical vein endothelial cells from amiodarone-induced damage by intermedin through activation of wnt/β-catenin signaling pathway, *Oxid. Med. Cell. Longev.* 2021 (2021), 8889408.
- [73] N. Pechlivani, R.A. Ajjan, Thrombosis and vascular inflammation in diabetes: mechanisms and potential therapeutic targets, *Front. Cardiovasc. Med.* 5 (2018).
- [74] G. Piazza, et al., Venous thromboembolism in patients with diabetes mellitus, *Am. J. Med.* 125 (2012) 709–716.
- [75] M. Abiola, et al., Activation of wnt/β-catenin signaling increases insulin sensitivity through a reciprocal regulation of Wnt10b and SREBP-1c in skeletal muscle cells, *PLoS One* 4 (2009), e8509.
- [76] R. Oughtred, et al., The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions, *Protein Sci. Publ. Protein Soc.* 30 (2021) 187–200.
- [77] D.A. Zarin, T. Tse, R.J. Williams, R.M. Califff, N.C. Ide, The ClinicalTrials.gov results database — update and key issues, *N. Engl. J. Med.* 364 (2011) 852–860.
- [78] J. Piñero González, et al., The DisGeNET Knowledge Platform for Disease Genomics: 2019 Update, 2020, <https://doi.org/10.1093/nar/gkz1021>.
- [79] S. Kerrien, et al., The IntAct molecular interaction database in 2012, *Nucleic Acids Res.* 40 (2012) D841–D846.
- [80] S.D. Harding, et al., The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY, *Nucleic Acids Res.* 46 (2018) D1091–D1106.
- [81] D. Domingo-Fernández, et al., Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): a web server for mechanism enrichment, *Bioinformatics* 33 (2017) 3679–3681.
- [82] N.P. Tattonetti, P.P. Ye, R. Daneshjou, R.B. Altman, Data-driven prediction of drug effects and interactions, *Sci. Transl. Med.* 4 (2012), 125ra31.
- [83] D. Ochoa, et al., Open Targets Platform: supporting systematic drug-target identification and prioritisation, *Nucleic Acids Res.* 49 (2021) D1302–D1310.
- [84] E.G. Cerami, et al., Pathway Commons, a web resource for biological pathway data, *Nucleic Acids Res.* 39 (2011) D685–D690.
- [85] J.C. Denny, et al., PheWAS: demonstrating the feasibility of a genome-wide scan to discover gene-disease associations, *Bioinformatics* 26 (2010) 1205–1210.
- [86] B. Jassal, et al., The reactome pathway knowledgebase, *Nucleic Acids Res.* 48 (2020) D498–D503.
- [87] M. Kuhn, I. Letunic, L.J. Jensen, P. Bork, The SIDER database of drugs and side effects, *Nucleic Acids Res.* 44 (2016) D1075–D1079.
- [88] M. Kanehisa, S.K.E.G.G. Goto, Kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (2000) 27–30.
- [89] Q. Duan, et al., LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures, *Nucleic Acids Res.* 42 (2014) W449–W460.
- [90] D. Himmelstein, L. Brueggeman, S. Baranzini, Consensus Signatures for LINCS L1000 Perturbations, 2016.
- [91] S. Schreiber, Cell Painting Morphological Profiling Assay, 2014.
- [92] D. Rogers, M. Hahn, Extended-connectivity fingerprints, *J. Chem. Inf. Model.* 50 (2010) 742–754.
- [93] G. Landrum, RDKit: Open-Source Cheminformatics, 2010.
- [94] A. Rives, et al., Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences, *Proc. Natl. Acad. Sci. U.S.A.* 118 (2021).
- [95] M. Ashburner, et al., Gene Ontology: tool for the unification of biology, *Nat. Genet.* 25 (2000) 25–29.
- [96] Gene Ontology Consortium, The Gene Ontology resource: enriching a GOld mine, *Nucleic Acids Res.* 49 (2021) D325–D334.
- [97] A.L. Beam, et al., Clinical concept embeddings learned from massive sources of multimodal medical data, *Pac. Symp. Biocomput. Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 25 (2020) 295–306.
- [98] T. Mikolov, M. Karafiat, L. Burget, J. Černocký, S. Khudanpur, Recurrent neural network based language model, *Proc. Interspeech 2010* (2010) 1045–1048, <https://doi.org/10.21437/Interspeech.2010-343>.
- [99] A. Lemsara, S. Ouadfel, H. Fröhlich, PathME: pathway based multi-modal sparse autoencoders for clustering of patient-level multi-omics data, *BMC Bioinf.* 21 (2020) 146.
- [100] M. Schlichtkrull, et al., Modeling relational data with graph convolutional networks, in: A. Gangemi, et al. (Eds.), *The Semantic Web*, Springer International Publishing, 2018, https://doi.org/10.1007/978-3-319-93417-4_38, 593–607.
- [101] D. Busbridge, D. Sherburn, P. Cavallo, N.Y. Hammerla, Relational Graph Attention Networks, 2019, <https://doi.org/10.48550/arxiv.1904.05811> arXiv.
- [102] M. Wang, et al., Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks, 2019 arXiv e-prints, <https://ui.adsabs.harvard.edu/abs/2019arXiv190901315W>.
- [103] A. Paszke, et al., PyTorch: an imperative style, high-performance deep learning library, in: *Advances in Neural Information Processing Systems Vol. 32* (Curran Associates, Inc., 2019).
- [104] W. Falcon, PyTorch Lightning, 2019.
- [105] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: a next-generation hyperparameter optimization framework, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19* 2623–2631, ACM Press, 2019, <https://doi.org/10.1145/3292500.3330701>.
- [106] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, in: *25th Annual Conference on Neural Information Processing Systems vol. 24, NIPS, 2011, 2011*.
- [107] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar, Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization, 2016, <https://doi.org/10.48550/arxiv.1603.06560> arXiv.
- [108] W. Hu, et al., Open Graph Benchmark: Datasets for Machine Learning on Graphs, 2020, <https://doi.org/10.48550/arxiv.2005.00687> arXiv.
- [109] N. Kokhlikyan, et al., Captum: A Unified and Generic Model Interpretability Library for PyTorch, 2020, <https://doi.org/10.48550/arxiv.2009.07896> arXiv.
- [110] Y. Zhou, et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, *Nat. Commun.* 10 (2019) 1523.
- [111] Y. Hochberg, Y. Benjamini, More powerful procedures for multiple significance testing, *Stat. Med.* 9 (1990) 811–818.