Graph Neural Network-Based Approaches to Drug Repurposing: A Comprehensive Survey

Alireza A.Tabatabaei
Faculty of Computer Engineering
University of Isfahan
Isfahan, Iran
Tabatabaei1999@outlook.com

Mohammad Ebrahim Mahdavi
Faculty of Computer Engineering
University of Isfahan
Isfahan, Iran
mohammad.mdv@mehr.ui.ac.ir

Ehsan Beiranvand
Faculty of Pharmacy
Shahid Beheshti University of Medical Sciences
Tehran, Iran
Beiranvand1999@gmail.com

Sajjad Gharaghani *
Laboratory of Bioinformatics and Drug Design (LBD)
Institute of Biochemistry and Biophysics
University of Tehran
Tehran, Iran
s.gharaghani@ut.ac.ir
* Corresponding author

Peyman Adibi*
Faculty of Computer Engineering
University of Isfahan
Isfahan, Iran
adibi@eng.ui.ac.ir
* Corresponding author

Abstract

Drug development is time-consuming and costly, with many efforts failing. About 90% of new molecules synthesized never reach the market, most of which have already undergone laboratory and animal testing phases. Drug repurposing, the use of drugs that successfully passed clinical trials for non-original purposes, reduces drug development time and cost. It's crucial for rare diseases and emerging diseases like COVID-19. Rare diseases lack sufficient incentives due to limited patient populations while emerging diseases require urgent drug development.

Various approaches exist for drug repurposing, among which researchers highly favor computational methods due to their systematic approach and the elimination of laboratory costs. One of the leading approaches within computational methods is the application of artificial intelligence (AI). In these methods, AI models propose drugs for new therapeutic targets. Subsequently, these proposed drugs undergo evaluation through in silico, in vitro, in vivo studies, and ultimately clinical trials.

Among AI-based methods, Graph Neural Networks (GNNs) have gained significant attention from researchers due to their ability to model problems in graph structures, which is highly suitable for drug repurposing tasks. This survey focuses on reviewing GNN-based approaches for drug repurposing. It provides a taxonomy of these models and examines various aspects such as model evaluation methods, implementation strategies, and datasets utilized in the field.

Keywords: Computational drug design, Drug repurposing, Drug repositioning, Graph neural networks, Artificial intelligence

1 Introduction

According to the US Food and Drug Administration (FDA), the development of new drugs follows five steps: (1) discovery and preclinical research, (2) safety studies, (3) clinical studies, (4) FDA review and approval, and (5) post-market safety monitoring. In the initial step, drug molecules are designed or selected using in-silico methods, which involve simulating the drug's effects using computational models. This step continues with in-vitro studies, which involve investigating the drug's effects on cells and tissues, followed by in-vivo studies, where the drug is tested on animal models. Finally, clinical studies are conducted to test the drug on human subjects.

Usually, the time required to complete these five steps is substantial. This process takes an average of 10 to 15 years. Additionally, many research efforts do not result in successful drug production, with a success rate of approximately 2.01% [1], [2]. The financial burden is also significant, with the development of a new drug molecule costing pharmaceutical companies about $12 million [3], as estimated in 2015. These challenges have shifted attention toward drug repurposing, which involves finding new applications for existing drugs. Also known as redirecting, repositioning, or reprofiling, drug repurposing offers a potentially faster and less expensive alternative to traditional drug development [4].

Today, finding or designing a lead compound, which is a potential medicinal molecule, is done in the in-silico environment based on Computer-Aided Drug Discovery (CADD) methods. CADD leverages computational techniques to save time and reduce costs associated with traditional methods, increasing the probability of success [5]. For example, anti-HIV drugs such as Atazanavir [6] and Saquinavir [7] were developed using CADD. Two prominent approaches within CADD are Structure-Based Drug Discovery (SBDD) and Ligand-Based Drug Discovery (LBDD) [8], [9].

Candidates for drug repurposing are typically selected from two sources: (1) existing market drugs and (2) drugs that failed clinical efficacy trials for reasons unrelated to drug safety. The advantage of drug repurposing is that the safety profile of these drugs is already well-established, which reduces the cost and duration of clinical studies [10]. For instance, Colchicine, originally used for Gout, was approved by the FDA in 2023 to reduce the risk of Atherosclerosis [11]. Similarly, Sildenafil, initially used for pulmonary hypertension, was repurposed to treat erectile dysfunction [10]. Drug repurposing can also expedite the development of treatments for orphan diseases and pandemics, such as COVID-19, where rapid research and development are crucial [13], [14]. Additionally, drug repurposing can be utilized as a tool to discover antimicrobial properties in non-antibiotic drugs and can be proposed as a solution to the problem of antibiotic resistance [14].

Historically, drug repurposing was often serendipitous or based on clinical observations, lacking a systematic approach. Today, systematic methods such as phenotypic screening and bioinformatics are more logical and rational, offering greater opportunities for success [10]. Approximately 30% of existing drugs approved by the FDA originate from drug repurposing, with even higher percentages in certain diseases, such as 60% for leishmaniasis disease treatments [15].

To provide a history of computational methods, we must refer back to the famous statement by Galileo Galilei, who, in the 17th century, asserted that God wrote the universe in the language of mathematics. Following this proclamation, computation-based approaches for studying the surrounding world

intensified, applying mathematical laws to natural phenomena. Examples of this approach include the emergence of modern physics and chemistry. However, as we know, due to the extreme complexity of models and the large volume of computations, applying mathematical models to biological phenomena has presented significant challenges. In recent years, remarkable advancements in computer systems, alongside the increased processing power and the development of soft computing and artificial intelligence methods, have significantly expanded the field of computational drug design and, more specifically, computational drug repurposing. Here, we will review the computational drug repurposing, which its hierarchical diagram is illustrated in Figure 1.

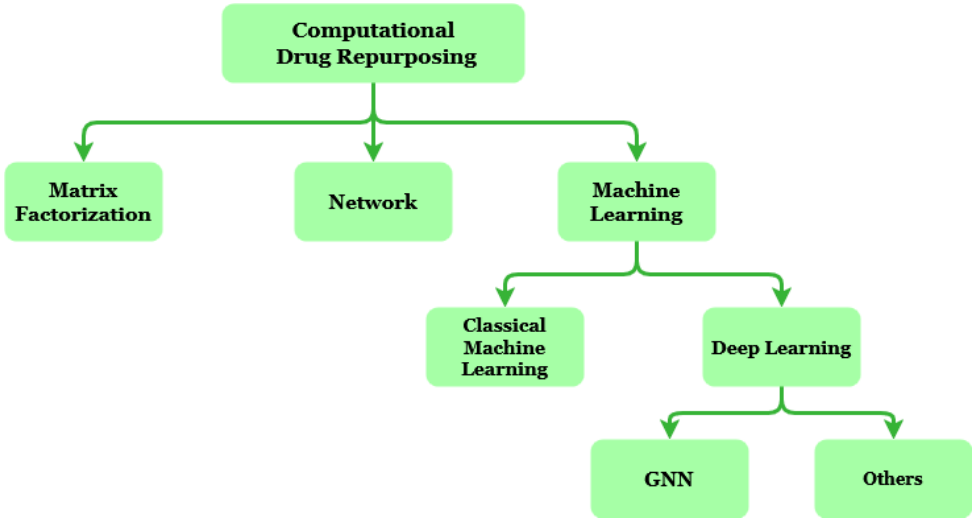

Figure 1 Computational drug repurposing methods

The computational methods employed for drug repurposing, shown in the first level of hierarchy of Figure 1, are described as follows:

1. *Matrix factorization-based approaches*: Matrix factorization is a linear algebra technique used to decompose a matrix into the product of two or more matrices. It is used in many recommender systems as collaborative filtering, and many researchers have utilized this technique in studies such as [16] [17] [18] [19] [20] for drug repurposing.

2. *Network-based approaches*: A network is a vast collection of nodes interconnected by links. Consequently, many referred to networks as graphs until advances in computer systems have drawn researchers' attention to structures such as social and business networks. They became interested in examining the common properties of these networks regardless of their content [21], with these investigations starting with publications such as [22] and [23]. Thus, network analysis involves studying complex networks to understand their shared properties without focusing on their specific nature. Given that large networks can include biological data and are inherently complex [24], some researchers have turned to network analysis for studying complex biological networks [25], including drug repurposing, as discussed in [26].

3. *Machine learning-based approaches*: The goal of machine learning is to design algorithms capable of building models from data to make predictions for new data or, in other words, to learn patterns within the data [27]. Consequently, many algorithms have been designed, each varying in complexity and application based on the specific problem. We classify machine learning methods for computational drug repurposing into two categories:

3.a. Classical machine learning approaches: We use the term classic machine learning to refer to machine learning methods that do not involve deep learning. This category includes models such as logistic regression, support vector machines (SVM) [28], and random forests [29]. For example, [30] used logistic regression, [31] applied SVM, and in [32] random forest has been used to predict drug-disease associations.

3.b. Deep learning approaches: Deep learning generally refers to a subset of machine learning algorithms comprising multiple layers capable of modeling complex data structures and relationships [27]. The fundamental model in deep learning is the multi-layer perceptron (MLP), which consists of one or more neurons per layer. These neurons are inspired by the biological neurons found in the human nervous system. It has been proven that MLPs satisfy the universal approximation theorem, which states that MLPs can estimate any function or pattern present in the data. Originally, MLPs were designed for tabular data, so extensive research has expanded deep learning models for other data types. For example, convolutional neural networks (CNNs) were developed for image data, as images are represented as matrices in computers, allowing CNNs to process them efficiently. For sequential data, such as text, deep learning models like Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) were introduced [33]. After these initial models were developed, attention mechanisms inspired by human attention processes were introduced. Initially designed for machine translation, attention mechanisms quickly found applications in various tasks, including computer vision and speech processing. With the evolution of attention, the Transformer model was introduced, marking a breakthrough in deep learning. This model leveraged attention mechanisms across a broad spectrum of tasks, from natural language processing (NLP) to computer vision, significantly advancing performance and versatility in these fields [34].

Graph Neural Networks (GNNs) are deep learning models designed to comprehend data with graph structures efficiently [33]. These models aim to achieve high accuracy in graph-related tasks, such as node classification and edge prediction. Techniques such as Graph Attention Networks (GAT), Graph Convolutional Networks (GCN), and graph transformers are notable examples in this field Many research applied [35].

This article reviews drug repurposing studies based on GNNs. To the best of our knowledge, it is the first survey in this field. It is structured as follows. Section 2 presents a new taxonomy and reviews the models. In Section 3, common datasets and implementations are introduced. Section 4 is dedicated to evaluation techniques. Section 5 gives discussion, where a new methodology for GNN-based drug repurposing is suggested. Finally, Section 6 presents concluding remarks.

2 Taxonomy

In this article, we aim to present various drug repurposing models that leverage GNNs, categorizing them according to the specific methodologies employed. As illustrated in Figure 2, the classification encompasses different approaches and techniques utilized in these models. Through this structured presentation, we seek to provide a comprehensive understanding of how GNN-based models are applied in drug repurposing, highlighting each method's strengths and unique features.

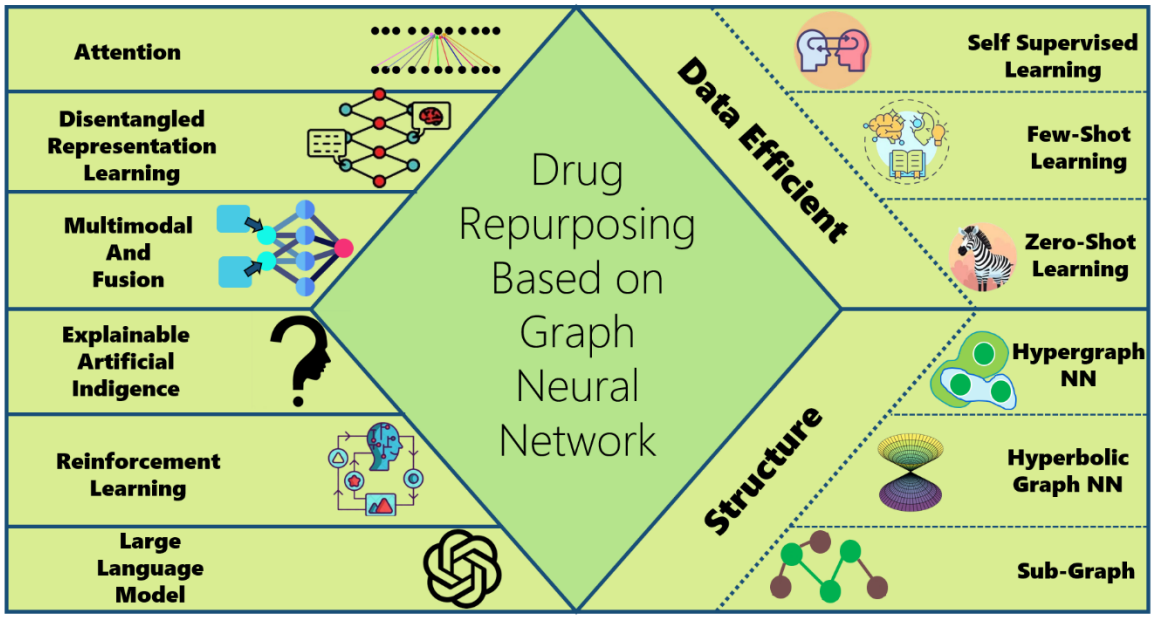

Figure 2. Taxonomy of GNNs for drug repurposing models

2.1. Attention based methods

Attention mechanisms, inspired by human cognitive processes, enable models to dynamically focus on the most relevant parts of input data, thereby enhancing both interpretability and performance. Initially developed for machine translation, attention mechanisms have since been applied to a wide range of AI-based applications, including computer vision [34]. Over time, these mechanisms have been integrated into GNNs, leading to the creation of GATs [36]. GATs enhance GNNs by allowing them to learn to assign different levels of attention to different edges in a graph, thereby improving the model's ability to capture important relationships within the data. Another prominent attention-based neural network architecture is the transformer [37], originally introduced in the context of NLP. Transformers have revolutionized NLP by enabling models to process entire sequences of words at once, rather than sequentially, and to focus attention on the most relevant parts of the input sequence. The transformer model's success has led to its adoption in various other AI applications, including computer vision and

reinforcement learning. The principles of transformers have also been applied to GNNs, resulting in the development of Graph Transformer Networks (GTNs) [38].GTNs leverage the transformer architecture to enhance the capability of GNNs in capturing complex dependencies and interactions within graph-structured data.

In this context DRGBCN model [39], employs three distinct graphs: drug-drug, disease-disease, and drug-disease. The DRGBCN model is composed of three primary components. First, the separate feature encoder processes drug and disease embeddings independently by leveraging the drug-drug similarity network and the disease-disease similarity network. Using a GCN, the encoder extracts detailed embeddings, which are then integrated using a layer attention mechanism. This process generates comprehensive representations of drugs and diseases. Second, the bilinear attention network captures complex relationships between drug and disease feature representations. Third, the decoder, an MLP, translates these embeddings into the predictions.

Another model is MGATRx [40] , which constructs a heterogeneous graph from various datasets. This model leverages diverse biomedical data types, encoding the information using multiview linear transformation layers and GATs. The multiview transformation layers facilitate the integration of different feature types, while GATs apply attention mechanism to identify the most relevant parts of the graph. The final step involves a single linear layer for multi-label classification.

The [41] address the COVID-19 treatment, it proposes a mixed graph network model that integrates attention mechanism with GCNs to predict drug-target affinity for SARS-CoV-2 proteins. The process begins with the extraction of atomic features and the generation of an undirected graph using the RDKit tool [42]. Protein sequences are encoded, and their features are extracted using CNNs. The study's key contributions include the use of graph representation learning with attention mechanisms for drug feature extraction, the construction of a mixed graph network model to predict drug-target affinity, and the identification of seven drugs that act on SARS-CoV-2 target proteins. By converting SMILES strings to molecular graphs and employing deep learning algorithms, the model effectively learns drug representations. The combination of GCN and GAT enhances feature extraction, while CNN layers process protein sequences. The resulting representations are cascaded through fully connected layers to predict drug-target affinity.

The Relgraph [43] model, utilizes meta-nodes which are categorized into entity-meta-nodes and predicate meta-nodes. The entity-meta-nodes capture the influence of entities on predicates, and vice versa for predicate-meta-nodes. The authors define "entity graphs" for each relation in the knowledge graph (KG), containing entity-meta-nodes, and "relation graphs" for each entity, containing predicate-meta-nodes, interconnected by "entity links." A "truth function" is introduced to determine the existence of facts in the KG based on these connections. The Relgraph attention network uses attention mechanisms to create well-trained embeddings for knowledge graph completion (KGC) [44] and rule mining. It employs two interconnected GATs that share predicate embeddings: the entity GAT focuses on message passing and attention among entity-meta-nodes, while the relation GAT addresses attention among predicate-meta-nodes.

DRAW [45] builds networks of drugs and diseases using chemical, therapeutic, and phenotype similarities. It extracts subgraphs and labels nodes based on their proximity to targets. Convolutional

networks process these subgraphs, while attention mechanisms generate node profiles. A perceptron then predicts drug-disease links. GTDDA [46] merges convolutional networks with graph transformers to map drugs and diseases. It analyzes similarity networks, replacing standard attention with graph-specific attention. A random forest algorithm then classifies the drug-disease pairs, reducing overfitting in predictions. WIGRL [47] creates networks depicting drug and disease similarities alongside their associations. It processes these networks using convolutional and attention mechanisms to extract node features. A perceptron then computes association scores, identifying potential drug repurposing candidates. DRTerHGAT [48] forms a three-way graph linking drugs, proteins, and diseases. It extracts protein features using a language model and compresses them with an Auto-encoder. The model processes homogeneous connections via GCN and heterogeneous links through attention networks, forecasting new drug-disease links. HGTDR [49] is an end-to-end model that employs heterogeneous graph transformers [50] on PrimeKG [51] for drug repurposing. The model utilizes ChemBERTa [52] and BioBERT [53] for initial embeddings, and ultimately, a fully connected neural network is used for predicting drug repurposing.

2.2. Disentangled representation learning based methods

Graph representation learning aims to embed graph-structured data into low-dimensional, continuous vector spaces while preserving the graph's inherent properties and relationships. This transformation facilitates the efficient analysis and application of machine learning algorithms on graphs, which enhances the performance of these algorithms [35]. A systematic neural network-based method for improving representation is disentangled representation learning. This approach seeks to develop models that can identify and separate underlying factors hidden in visible data into distinct and interpretable representations. Disentangled representation learning aims to separate distinct, independent, and informative factors that generate changes in the data. Single latent variables are sensitive to single underlying generative factors while remaining relatively stable to changes in other factors [54]. Among the widely used neural networks for disentangled representation learning is the family of Auto-encoders. Auto-encoders are neural networks consisting of an encoder and a decoder, providing not only low-dimensional representation but also a manifold of data [55], [56]. A special kind of auto-encoder is Variational Auto-Encoders (VAEs) [57], and the group of VAEs that receive graphs as input is called Variational Graph Auto-Encoders (VGAEs) [58].

The VGAEDR [59] model operates as follows: First, a drug-disease heterogeneous network is established based on three drug attributes, disease semantic information, and known drug-disease associations. Second, low-dimensional feature representations for heterogeneous networks are learned through a variational graph Auto-encoder module and a multi-layer convolutional module. Finally, the feature representation is fed to a fully connected layer and a SoftMax layer to predict new drug-disease associations.

Another model in this category [60] is a variational graph auto-encoder with multi-relational edges for node embedding, followed by a separate ranking model for drug prioritization. The VGAE takes the COVID-19 knowledge graph as input, where nodes represent entities (e.g. drugs, genes, phenotypes) and edges represent relationships between them (e.g. drug-target interactions, gene-pathway memberships). Another model is DRGCC [61] that combines clustering constraints and matrix factorization and then use GraphSAGE to get better representation.

The DDAGTP [62] model exemplifies the potential of integrating heterogeneous information networks for drug repurposing. By combining drug-disease associations, drug-protein associations, and protein-disease associations, the model forms a comprehensive heterogeneous information network that encapsulates the multifaceted relationships within biomedical data. The graph representation learning process begins with initial feature extraction via a Graph Auto-encoder (GAE), which reduces features to 64 dimensions. These features are then transformed using a graph transition probability matrix, calculated through node degrees and diagonal matrix construction. This matrix refines the feature representations by capturing the probability transitions between nodes. An improved convolutional Auto-encoder further enhances these features, ultimately aiding in the prediction of drug-disease associations using the CatBoost classifier [63]. The RAFGAE model [64] employs multilayer GATs and residual networks. It utilizes two Graph Auto-Encoders (GAEs) specifically for drugs and diseases and incorporates a free multiscale adversarial training technique to enhance the quality of the features. DRAGON [65] is an encoder-decoder model that utilizes drugs and proteins as features. The encoder architecture employs GraphSAGE, while drug features are derived using the DSMR model [66], and protein features are obtained using the [67] model.

2.3. Multimodal and data fusion approach

In many problems, to achieve better results, it is necessary to incorporate the information obtained from the various data sources into the model. When the data types are different, this approach is referred to as multimodal, and when the data types are the same, it is called data fusion [68].

One model in this domain is the DRHGCN [69], which addresses both intra-domain and inter-domain feature extraction to improve drug repurposing predictions. For intra-domain feature extraction, separate graph convolutional networks are used on drug-drug and disease-disease similarity networks, capturing internal structures within each domain. For inter-domain feature extraction, a combination of bilinear and traditional GCN aggregators models interactions between drugs and diseases. Embeddings from these processes are fused and refined with a layer attention mechanism, enhancing representation quality.

The BiFusion model [70] operates within an encoder-decoder architecture, with bipartite GCN layers learning embedding representations from a heterogeneous network. The MLP decoder reconstructs the drug-disease association matrix using these embeddings. GraphPK [71] is another model that improves in silico drug repositioning by integrating prior knowledge from knowledge graphs. It constructs a knowledge graph embedding various bio-entities and their interactions using TransD [72] , extracting meaningful prior knowledge. Feature extraction includes biological features from SMILES representations [68], known association-based features from a drug-disease bipartite graph using LightGCN [74], and knowledge graph-based features from TransD embeddings. These features are concatenated and processed through an MLP, optimizing drug-disease predictions.

The GCMM model [75] employs a multimodal attention mechanism within a GCN framework for drug repurposing. It constructs a heterogeneous network using chemical, therapeutic, semantic, and target-based similarities. The GCN encoder processes these similarity networks, and the multimodal attention mechanism assigns weights to multi-source features. A fully connected feature extractor integrates the information, and a matrix completion decoder predicts drug-disease associations, minimizing the Frobenius norm between predicted and actual matrices. The FuHLDR model [76] represents a multimodal

framework aimed at constructing attribute matrices for drugs, diseases, and proteins from various databases, such as molecular fingerprints and semantic similarities. It uses a GCN for lower-order representation learning, aggregating information from neighboring nodes in the heterogeneous information network. Higher-order representations are learned using a meta-path-based random walk strategy in metapath2vec model [72] capturing indirect relationships. These representations are integrated and used as input for an RVFL classifier[78] to predict new drug-disease associations.

The MilGNet model [79] projects different types of node features into the same dimensions using a HeteroGCN [80] to learn node representations from drug-drug and disease-disease graphs. It generates pseudo-meta-paths linking drugs and diseases through their neighbors, aggregating instance embeddings with the BiTrans approach. The model employs an attention mechanism to weigh and aggregate meta-path instances, predicting association probabilities for interpretable outputs. MilGNet's multi-instance learning approach allows for nuanced and interpretable predictions in drug repurposing.

STRGNN [81] integrates deep learning within multimodal networks using a three-part architecture. The attribute encoder converts information about individual entities (drugs, diseases, proteins) into numerical representations, using methods like fingerprint vectors for drugs and CNN and FNN (feed forward neural network) for proteins. The network encoder processes the network structure with a GNN. The decoder multiplies the feature vectors of drug and disease entities and applies a sigmoid function to determine the likelihood of a link between them. A core novelty of STRGNN is the incorporation of topological regularization, which addresses challenges in large-scale multimodal networks, enhancing the model's ability to manage complex topologies and improve prediction accuracy.

DRMAHGC [82] combines multiple similarity networks for drugs, diseases, and proteins. It uses a masked Auto-encoder for feature extraction and contrasts different graph views. The model synthesizes negative samples from positive ones and predicts associations using a multi-layer perceptron. TGCNDR [83] constructs a three-part network connecting drugs to diseases, proteins, and side effects. It uses shared drug nodes across networks to transfer information. Convolutional networks and attention mechanisms process the graph, while a skip-gram model predicts new drug-disease associations.

2.4. Structural approaches

Sometimes, changing the perspective on the nature of graphs or graph neural networks can be a useful approach to solving problems. This section introduces approaches based on hyperbolic graph neural networks, which involve transitioning the problem space from Euclidean to non-Euclidean. Following this, approaches based on hypergraphs are discussed, which extends the traditional notion of graphs. Finally, approaches based on the decomposition of graphs into smaller subgraphs are examined.

2.4.1. Hyperbolic GNN-based methods

In geometry, spaces are divided into two categories: Euclidean and non-Euclidean. Euclidean spaces are those in which two parallel lines never intersect, whereas non-Euclidean spaces are those in which parallel lines diverge from or converge to each other. The definitions of Euclidean and non-Euclidean geometry can be explained through the concept of curvature. Curvature describes how a geometric object interacts with a plane in space. (Figure 3) In Euclidean geometry, the curvature is zero, while in non-

Euclidean geometry, the curvature is a non-zero value. If the curvature is negative, we say that the space is hyperbolic. This concept of curvature can be extended to graphs. If the curvature of a graph is a non-zero value, it can be said that the graph exists in a non-Euclidean space. For instance, graphs such as trees exist in non-Euclidean space. It is also known that GNNs are defined for Euclidean spaces. For example, GCN are Euclidean functions. In contrast, for hyperbolic non-Euclidean spaces, we encounter the concept of Hyperbolic GNNs [84], witch specialized type of neural networks designed to operate on data represented in hyperbolic spaces, a non-Euclidean geometric framework that effectively captures hierarchical and complex relationships. These networks leverage the unique properties of hyperbolic space to efficiently model and embed data with hierarchical structures. This approach enables more compact and meaningful representations, leading to improved performance in tasks like node classification, link prediction, and clustering, particularly in scenarios where data naturally exhibits hierarchical patterns. For example, in hyperbolic graph neural networks, functions such as GCNs have been adapted to the characteristics of hyperbolic space [85].

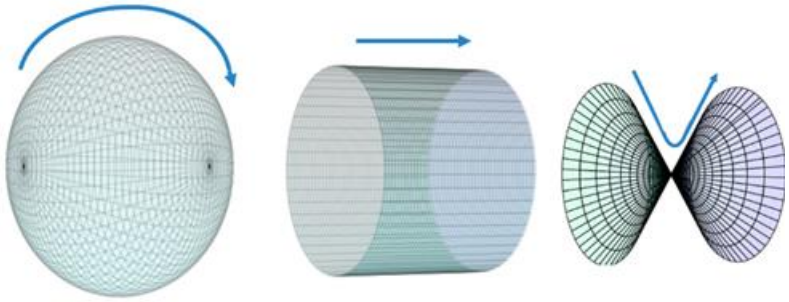

Figure 3. In the first shape we can observe positive curvature, in second we have zero curvature and in the last we can see negative curvature.

The HGNN-DR [15] model employs hyperbolic GNNs for drug repurposing, specifically targeting leishmaniasis disease. Leishmaniasis, a parasitic disease, presents a unique challenge compared to other diseases because the therapeutic target is not a human cell but the parasite itself. This distinction necessitates the creation of a graph comprising proteins related to both the disease and potential drugs. A critical aspect of HGNNs is neighborhood aggregation, which updates node representations based on information from their 1-hop neighbors. The HGNN-DR model extends previous HGNN approaches to address performance degradation during multi-step aggregation. This is achieved by decoupling feature transformation and neighborhood aggregation, thus enhancing the model's efficiency and accuracy in capturing the complex interactions within the graph.

2.4.2. Hypergraph-based methods

In graph theory, a graph is defined by the relationships between nodes through edges, with each edge connecting at most two nodes. A hypergraph generalizes this concept by introducing the notion of a hyperedge, which can connect any number of nodes (as illustrated in Figure 4) [86]. Hypergraphs have proven effective in modeling high-order interactions, which involve relationships among more than two elements. Given that many biological mechanisms result from the interactions of more than two

components, hypergraphs present a suitable option for modeling such complex systems. Hypergraph Neural Networks are a class of neural networks designed to learn the structure of hypergraphs.

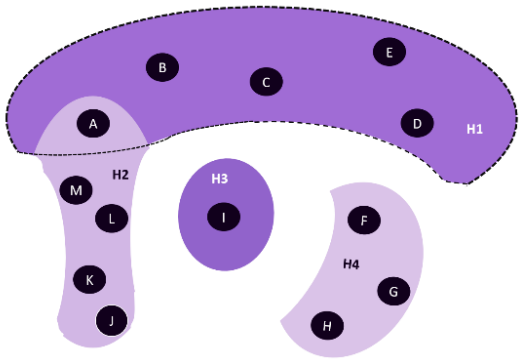

Figure 4. An example of a hypergraph. In this example we have 13 nodes and 4 hyperedge. Here node A is common between H1 and H2.

Hypergraphs have been employed for drug repurposing [87]. In this method, subgraphs related to Alzheimer's disease are initially generated. Hyperedges are then produced based on biological pathways, and a graph is constructed from the hypergraph, where each node represents a biological pathway, and the edges signify connections between these pathways. Subsequently, the node2vec [88] algorithm is applied to this graph to generate meaningful embeddings for further analysis. This approach leverages the ability of hypergraphs to capture multifaceted interactions, thereby enhancing the process of drug repurposing by providing a more nuanced understanding of the biological underpinnings of diseases.

In the HGDD model [89], two initial graphs are utilized: the drug-disease graph and the drug-side effects graph. Following this, the node2vec algorithm is applied to extract features, which are then used in the construction of hypergraphs. The HGDD paper details the creation of two specific hypergraphs. The first hypergraph consists of hyperedges where each hyperedge connects drugs that target a common disease (i.e., drugs and common disease in each hyperedge). The second hypergraph comprises hyperedges where each hyperedge includes diseases that are treated by a common drug. Hyper GCN are then employed to extract information from these hypergraphs, enabling the identification of new potential diseases that a given drug can treat.

Another model is EMPHCN [90], Whose innovation lies in its dual message-passing mechanism, which enhances the extraction of intra-domain and inter-domain embeddings of drugs and diseases. Intra-domain message passing is achieved through hypergraph convolutional networks, capturing intricate drug and disease characteristics within their respective domains. Meanwhile, inter-domain message passing employs a combination of node and edge embeddings using GCN, facilitating the integration of known drug-disease associations with additional protein interaction data. This combination allows EMPHCN to effectively propagate information across different biological entities, leading to more accurate and explainable predictions of drug repositioning opportunities.

2.4.3. Subgraph-based methods

In some models, subgraphs extracted from the primary problem graph are utilized to identify suitable drugs for repurposing. The PSGCN model [91] attempts to solve the problem by converting link prediction into graph classification. Given that the mechanism of drug action varies across different diseases, it is essential for the embeddings to differ according to the specific drug and disease. Therefore, in this approach, subgraphs are initially generated from the primary graph, assisting in ensuring drug resistance under various conditions. These subgraphs are termed pattern-specific graphs. The PSGCN model comprises three components: a pattern-specific graph extractor module, a module for learning representations of pattern-specific graphs, and a link prediction module.

Another model in this category is CBR-SUBG [92], which analyzes subgraphs within a larger knowledge network to repurpose drugs. It encodes these subgraphs using neural networks to uncover potential drug action pathways. The model finds similar drug cases, extracts their reasoning paths, and applies them to new drugs, balancing prediction accuracy with biological interpretability.

2.5. Data efficient approaches

Data-efficient approaches are employed in situations where reliable data is deficient or flawed. To address this issue, techniques such as self-supervised learning and few-shot learning are utilized. These approaches have gained attention in the context of drug repurposing because the graphs commonly encountered in this field tend to have sparse structures, and the number of positive samples is significantly lower than that of negative ones. Additionally, the emergence of diseases with limited reliable data, such as COVID-19, has highlighted the importance of these methods.

2.5.1. Self-supervised learning-based methods

Self-supervised learning (SSL) is an unsupervised approach in which the system learns to generate its own labels from the data, eliminating the need for labeled datasets. This approach leverages the inherent structure within the data to create pseudo-labels for training. By exploiting the relationships and patterns present in the data throughout some procedures called pre-text tasks, SSL allows models to learn useful representations that can be fine-tuned for specific tasks called downstream tasks, such as image classification or natural language processing. This is particularly advantageous in scenarios where labeled data is scarce or expensive to obtain, as SSL can utilize vast amounts of unlabeled data to pre-train models effectively [93].

The SADR [94] method utilizes SSL for drug repurposing, by employing contrastive learning to address the noise present in the data, which arises from the recording samples during datasets creation process and the inherent sparsity of these datasets. The DRGCL [95] method also leverages contrastive learning to extract more refined global and semantic features from the graph. Additionally, the SGCD [96] model, which is specifically tailored for drug repurposing, demonstrates superior performance in comparison to SGL, owing to its customized application of contrastive learning.

2.5.2. Few-shot learning-based methods

Few-shot learning (FSL) refers to the machine learning scenarios where a model is trained to perform a task with very few training examples. This contrasts with traditional machine learning models, which typically require large amounts of labeled data to achieve high performance [97].

The MLAN [98] model filters entity neighborhoods using attention networks. It analyzes triplet relationships among entities and learns meta-representations for each relation type. The model adapts quickly to new tasks through meta-learning techniques, a usual way to handle few-shot problems, enhancing its versatility in drug repurposing scenarios.

FSL has been used for COVID-19 drug repurposing [99] to tackle the challenge of predicting interactions with limited examples. This is especially relevant for new diseases like COVID-19 with scarce known interactions between viral proteins and potential inhibitors. By leveraging prior knowledge and learning from a few instances, FSL enables models to generalize and make accurate predictions with minimal data, significantly reducing the time and cost of drug discovery compared to traditional methods. The proposed model, an inductive Relational Graph Convolutional Network (I-RGCN), extends the standard RGCN by incorporating inductive learning capabilities to learn embeddings for relations from node features. It uses a Multi-Layer Perceptron (MLP) to compute relation embeddings and excels in FSL scenarios. Applied to the Drug Repurposing Knowledge Graph (DRKG), the I-RGCN predicts potential drug-disease interactions, aiming to repurpose existing drugs for treating COVID-19. This inductive approach to relation embedding learning makes the I-RGCN particularly effective for addressing the challenges of drug repositioning in new diseases.

2.5.3. Zero-shot learning-based methods

Zero-shot learning (ZSL) is a type of learning that can detect samples of classes without any data sample seen in the training in the main domain, but with samples in other auxiliary domain. TXGNN [100] is a model designed for drug repurposing using ZSL, which is employed to develop a drug repurposing model specifically for diseases without established treatments. In addition to the graph-based geometric deep learning techniques, it also incorporates metric learning to enhance performance.

2.6. Explainable approaches

Explainable Artificial Intelligence (XAI) encompasses techniques and methods designed to render the decisions and behaviors of AI systems comprehensible to humans. The primary objective of XAI is to provide clear, human-readable explanations of how AI models reach their conclusions, thereby fostering trust and transparency in AI applications. This is particularly crucial in domains where decision-making must be interpretable, such as healthcare, finance, and legal systems. In these fields, stakeholders need to understand, validate, and challenge AI decisions. Moreover, XAI aids in identifying and mitigating biases, enhancing model robustness, and complying with regulatory requirements by making AI processes more transparent and accountable [101].

In this context, GraphIX [102], an explainable drug repositioning framework, is introduced. GraphIX leverages biological networks to achieve its objectives. The framework employs GNN to learn the

network weights and node features from a knowledge graph that includes disease, drug, and protein nodes. GraphIX distinguishes between these node types based on the graph structure, even in the absence of explicit node type information. The model predicts disease-drug associations and calculates the contribution values of nearby nodes, thereby identifying significant proteins for pharmacological understanding.

EDEN [103] states that because a lot of computational methods often lack interpretability, it is difficult to understand the underlying mechanisms of drug-disease associations. GNN is utilized by EDEN to capture both local semantics and global structures within a disease information network. By maximizing mutual information between these elements, interpretable embeddings that provide meaningful insights into why certain drugs may be effective for specific diseases can be created by the framework. One of the key innovations of EDEN is its ability to integrate various types of biomedical entities and relationships into a unified, low-dimensional space. This is achieved by aggregating neighboring messages based on different semantic connections, preserving the internal knowledge structure. The end-to-end learning process ensures that drug-disease associations are not only predicted by the model but also explained by highlighting relevant paths in the embedding space.

Another noteworthy method is iDPath [104], which offers explainable artificial intelligence solutions in the realm of drug repurposing. The framework utilizes GCN and LSTM networks to capture both the global connectivity and detailed mechanistic pathways of drugs. Additionally, two attention modules (path attention and node attention) are employed by iDPath to enhance the interpretability of its predictions. These modules identify critical paths between drugs and diseases, which are crucial not only for predicting drug efficacy but also for understanding the underlying biological mechanisms.

XG4Repo [105] innovates in drug repurposing by integrating explainable artificial intelligence with knowledge graphs to predict and explain potential new uses for existing drugs. Unlike traditional models, it focuses on interpretability, converting prediction rules into natural language and generating queries to trace paths in the Hetionet [106] graph, making explanations accessible to human experts.

2.7. Reinforcement learning-based methods

Reinforcement learning is a type of machine learning wherein an agent learns to make decisions by receiving rewards or penalties for actions taken within an environment. The agent aims to maximize cumulative rewards through trial-and-error interactions, thereby improving its performance over time[27].

The KGML-xDTD model [107] for drug repurposing and mechanism of action (MOA) prediction is structured into two key components. The first component, drug repurposing prediction, employs GraphSAGE [108] to generate node embeddings by analyzing the local neighborhoods of drugs and diseases within a knowledge graph. These embeddings are subsequently combined with PubMed BERT [109] to enhance classification accuracy. The integration of GraphSAGE and PubMed BERT leverages the structural and textual information, respectively, to provide a comprehensive feature representation. The second component, MOA Prediction, elucidates how a drug might work for a predicted treatable disease. This involves an actor network, a core reinforcement learning component that explores the knowledge graph by selecting the next node to visit based on the current state, aiming to reach the disease node. The actor-network's decisions are evaluated by a critic network, which estimates the potential future

rewards of its actions. Additionally, discriminator Networks guide the exploration toward biologically relevant paths, ensuring that the model's predictions are grounded in biological plausibility.

2.8. Large language models based approaches

Large Language Models (LLMs) are more complex neural networks based on transformer architecture, designed to process and generate human language. Their vast scale (characterized by billions of parameters) and training on massive datasets enable them to achieve state-of-the-art performance in NLP tasks, while also showing adaptability in various other domains [110].

Large Optimized Vector Embeddings Network or LOVENet [111] is an LLM-based drug repurposing approach. The architecture consists of three main components: Knowledge graph representation, LLM embedding, and fusion Layer for Prediction. In knowledge graph representation, A heterogeneous graph is constructed to model drug-disease relationships, integrating data on drug-drug, disease-disease, and drug-disease associations. This structured network is a foundation for understanding associations through graph-based embeddings generated by GNN. In LLM embedding, use Llama2 [112] to process textual prompts like “Is this drug related to this disease?” This generates embeddings that capture contextual and generative insights into drug-disease associations. In fusion layer for Prediction, the Outputs from the knowledge graph embeddings and the LLM embeddings are normalized and concatenated within a fusion layer. The resulting representation is passed through a fully connected network, enabling LOVENet to predict the probability of a relationship between drugs and diseases. This LOVENet claims that combining structured data from the knowledge graph and textual insights from Llama2 significantly enhances predictive performance.

LLM-DDA[113] is another model that leverages LLMs in drug repurposing. While GNNs have demonstrated strong performance in this field, they are limited by their dependence on graph topology, making them susceptible to noisy graphs. To address this issue, LLM-DDA integrates LLMs, recognizing their significant potential in reasoning. By combining the strengths of both GNNs and LLMs, the model aims to enhance drug repurposing efforts.

3 Common datasets and implementations

It is clear that the implementation of models is of great importance, and the data used as inputs to these models significantly impact their performance. In this section, we introduce the datasets and implementation methods that are utilized.

3.1 . Datasets

Datasets are integral to applying GNNs in drug repurposing, serving as the foundational element for constructing models that can predict new therapeutic uses for existing drugs. The importance of data in GNNs is underscored by the necessity for high-quality, diverse, and comprehensive datasets that capture the intricate relationships between drugs, their targets, and the resultant phenotypic effects. These datasets not only fuel the computational models but also ensure the validity and applicability of the predictions made by these models.

Drug repurposing involves identifying new medical uses for existing drugs, necessitating a robust understanding of various biological and chemical interactions. This process relies heavily on datasets that provide detailed information on drug-target interactions, chemical properties, genetic expressions, and disease pathways. These databases compile extensive data on small molecules, their biological targets, pathways, and the associated diseases, facilitating the identification of potential drug repurposing opportunities.

These resources enable researchers to access and integrate diverse data types, essential for developing predictive models using GNNs. A list of common datasets used in the field of drug repurposing is collected in Appendix 1.

3.2 . Implementation

Implementing models is of paramount importance, as it enables researchers not only to evaluate the designed models but also to utilize them practically. Various methods exist for model implementation; however, in the articles reviewed in this paper, the Python programming language and MATLAB have been used as implementation approaches. Appendix 2 includes links to the implemented codes of the articles, along with the implementation methods and libraries used.

Python as a high-level programming language is widely used for implementing artificial intelligence models, particularly GNNs, due to its simplicity compared to other programming languages, the availability of numerous libraries for AI programming, and its open-source and free nature. The libraries used in the reviewed articles include PyTorch [114], PyTorch Geometric [115], TensorFlow [116], DGL [117], and DeepSNAP [118]. As evident in Table 1, PyTorch is the most common library for implementing GNNs. Supported by Meta, this library offers robust support for tensor computations. PyTorch Geometric, built on top of PyTorch, provides common GNN models, data loaders, metrics, and datasets suitable for GNNs. TensorFlow, developed by Google, is a deep learning library that offers powerful tools for building and training neural networks and has recently made efforts to facilitate GNN development [119]. DGL, a library designed for GNNs, supports multiple deep learning frameworks, including PyTorch and TensorFlow, highlighting its compatibility with these libraries. DeepSNAP, developed by Stanford University, simplifies the integration of graph data with deep learning models.

MATLAB [120] is highly suitable for matrix computations and has the capability to train neural networks and GNNs [121]. However, compared to Python, it is less commonly used by developers.

4 Evaluation

Model evaluation is crucial, as it enables researchers to reconsider their data or model architecture if their designed model does not meet their objectives. This evaluation (as shown in Figure 5) occurs in two steps.

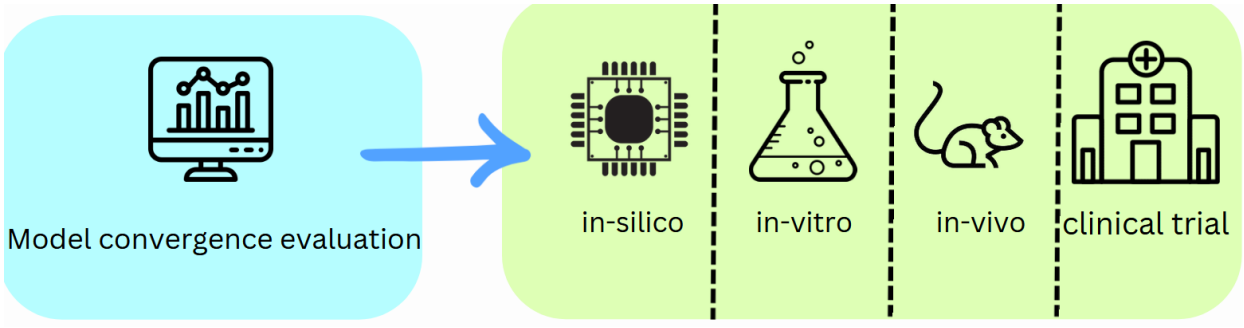

Figure 5. Evaluation steps

4.1. Model convergence evaluation

In the evaluation of graph neural network (GNN) models for drug repurposing, the primary focus is on measuring model convergence, which refers to the optimization of weights through training to minimize the loss function. This process involves using metrics like Area Under the Curve (AUC), Receiver Operating Characteristic (ROC) curve, Area Under the Precision-Recall Curve (AUPR), and F1-score to assess performance. A typical evaluation approach includes splitting the dataset into training and test sets, where the training set is used to adjust the model's weights, and the test set is used to validate its performance. K-fold cross-validation is often employed to ensure a more reliable evaluation by averaging results over multiple iterations, reducing bias, and improving the robustness of the performance estimates. By tracking metrics such as AUC and AUROC, researchers can ensure that the model generalizes well to new, unseen data.

A key evaluation tool in this process is the ROC curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) to visualize the trade-offs between sensitivity and specificity. The AUC metric, derived from the ROC curve, provides a quantitative measure of the model's ability to distinguish between positive and negative classes. This is particularly important in drug repurposing tasks, where datasets are often imbalanced, and traditional metrics like accuracy may be misleading. The ROC curve helps mitigate this issue by focusing on the performance across all thresholds. Additionally, an ablation study is often conducted after evaluation to understand the contribution of different components of the model. By systematically removing or altering model features, researchers can identify which elements are critical for achieving optimal performance, providing valuable insights for refining GNN models in drug repurposing tasks.

4.2. Recommended drug evaluation

However, the model's adequate performance in the evaluation with steps 1 and 2 criteria does not guarantee that the drugs proposed by this model are suitable for repurposing. In light of the multifaceted nature of the human body and the multitude of variables that contribute to its complexity, it is not a straightforward conclusion to ascertain that the top-n list is fully capable of drug repurposing. The re-evaluation steps of these molecules extracted by the model are analogous to evaluating a new molecule for the disease in question, except that safety studies are not conducted. It is noteworthy that safety studies are excluded if the proposed dose is near the maximum tolerated dose. Eliminating safety studies results in significant savings in time and financial resources. As mentioned in the introduction, the

following process is usually used to check the effectiveness of a drug: in-silico screening, in-vitro screening, in-vivo screening and clinical trial.

5 Discussion

This paper presented a comprehensive survey about different approaches of applying graph neural networks to drug repurposing problem. Accordingly, a method for constructing a graph-based model by analyzing existing literature is suggested. Although the method is grounded in scholarly references, specific challenges are identified, and recommendations are offered, supported with detailed examples. The process is illustrated in Figure 6, and is further discussed below.

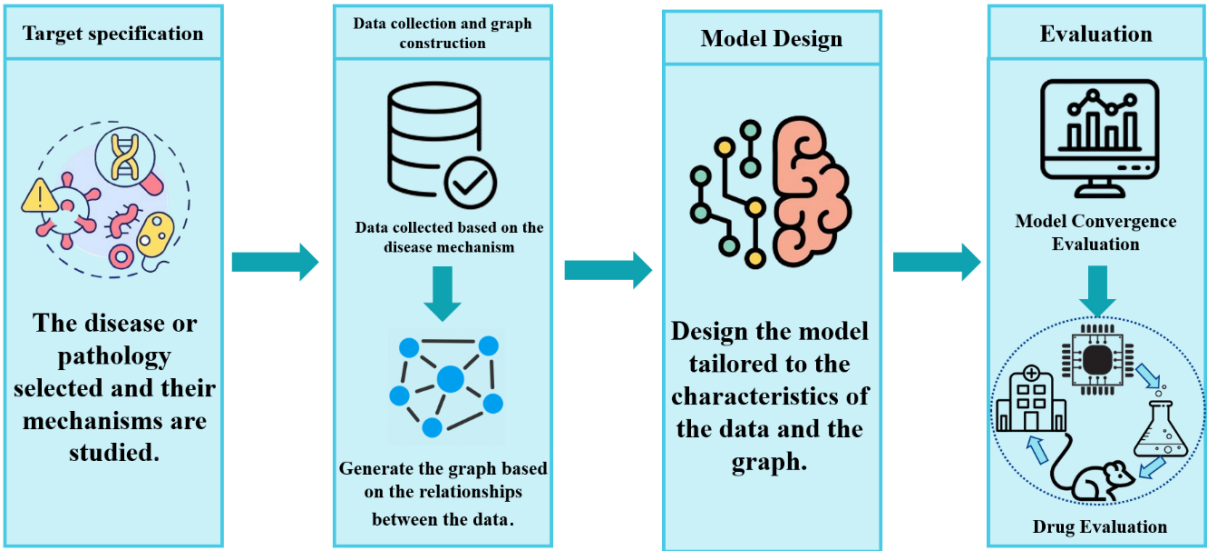

Figure 6. Graph neural network based drug repurposing developing method

5.1. Target specification: disease or pathology

In the initial step, it is essential to identify the disease for which drug repurposing will be pursued. Subsequently, the mechanisms involved in the pathology of the disease must be determined. This step is critical for effectively modeling the problem. An alternative approach at this step involves identifying common factors in the pathology of different diseases. If this approach is chosen, the resulting model will have the capability to repurpose drugs for similar pathologies across various diseases.

Incorporating biological knowledge into research is highly recommended to establish more precise and specific targets, thereby increasing the probability of achieving successful outcomes. For instance, in [122] is the utilization of PINK1's involvement in the pathogenesis of Parkinson's disease to guide drug repurposing efforts, facilitated through the development of a computational model or in [123], due to the importance of tissues in the mechanism of drug action, tissue-specific protein-protein interaction networks have been used.

It is important to note that some studies make minimal distinctions between drug target identification and drug repurposing. While these two concepts may overlap, it is recommended to design separate models for each. The term drug target identification refers to the discovery of new targets within the disease pathology for which drugs already exist or are yet to be developed. In contrast, drug repurposing refers to identifying a new application for a drug that has successfully passed various phases of human clinical trials.

5.2. Data collection and graph construction

In the previous step, the required data to be collected was identified. In this step, it is necessary to proceed with dataset collection, preprocessing, and filtering.

This step is highly significant, and any errors at this stage may adversely affect the overall outcome of the work. For instance, in [124] ,specific study on drug repurposing for COVID-19, excluded drugs from the categories D and X [125] during the data preprocessing stage, as these drugs are unsuitable for pregnant women. However, it is worth noting that such drugs might still be appropriate for treating COVID-19 in non-pregnant individuals. Overall, the negative impact of such preprocessing is more pronounced in tasks based on GNNs, as certain nodes can contribute to improving the final outcome solely through message passing.

At this stage, it is essential to examine the characteristics of the data, as certain datasets may include specific features that require consideration in subsequent stages. For example, some datasets contain a significant amount of predicted data (values estimated using statistical methods or artificial intelligence techniques). Researchers should take such features into account to ensure accurate analysis and modeling.

At this graph construction stage, the graph must be constructed based on the type of relationships between components identified in Step 1. A critical consideration in this step is selecting appropriate features for the graph's components. Choosing suitable features can significantly enhance the model's performance. For instance, in a specific study [49], BioBERT was utilized to generate the initial embeddings for diseases, while ChemBERTa was employed to generate initial embeddings for drugs. The authors emphasize that obtaining suitable initial embeddings for the graph components, tailored to the characteristics identified in Step 1, is highly advantageous.

In the following, we discuss the general concept of initial drug embeddings, which represent the most critical component in these studies. Although various drug-related features exist, the molecular structure of a drug holds undeniable significance. Typically, drug structures are represented in three formats: three-dimensional (3D), two-dimensional (2D), and SMILE (Simplified Molecular Input Line Entry System). The 3D structure of a drug contains information that cannot be fully extracted from its 2D representation. For example, as shown in Figure 7, Omeprazole has a chiral center, resulting in two enantiomers, R and S. The S-enantiomer is therapeutically more effective as a proton pump inhibitor due to its favorable pharmacokinetics and metabolic pathway, while the R-enantiomer exhibits reduced efficacy [126].

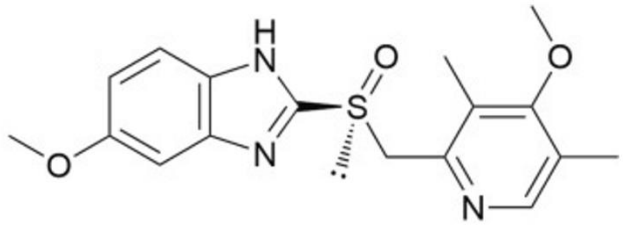

Figure 7. Omeprazole 2D structure. The 2D structures of S-OME and R-OME are represented in a way that they cannot be distinguished from each other in the 2D format.

Notably, enantiomers have distinct 3D structures despite sharing the same 2D connectivity. In some cases, inactive enantiomers may not only lack therapeutic properties but could also exhibit toxicity.

While the 3D structure provides comprehensive molecular information, embeddings derived from it are not necessarily richer than those obtained from the 2D structure. The richness of an embedding largely depends on the accuracy of the model generating it. The utility of embeddings is also influenced by the specific application, requiring researchers to select an appropriate model if the molecular structure is deemed relevant to their task.

Several models have been designed to extract drug embeddings from 3D structures, 2D structures, and SMILE representations. Models like Transformer-m [127] aim to leverage both 2D and 3D structures to generate embeddings. Additionally, alternative approaches for producing drug representations are discussed in [128]. Furthermore, drug embeddings can be enriched by incorporating other features such as pKa, pKb, solubility, lipophilicity. These additional features can enhance the quality of the initial representation and improve downstream performance

5.3. Model design

At this step, it is necessary to design an appropriate model based on the characteristics of the data, graph, and disease. The taxonomy section provides a comprehensive explanation of the methods used and the conditions under which each method is applicable. For example, in cases where the data is relatively limited, SSL-based methods can be a suitable choice.

5.4. Evaluation

In this section, it is essential to first evaluate the convergence and then assess the efficacy of the proposed drugs. The methods for these evaluations are discussed in detail in the evaluation section. Assessing drug efficacy is a critical step in research and must be conducted at one of the following levels: in-silico, in-vitro, or in-vivo. For instance, [129] presented a computational model for drug repurposing targeting the focal segmental glomerulosclerosis disease, demonstrating the efficacy of the proposed drug through in-vivo testing. In addition , [69] validated the proposed drug for Alzheimer's disease using molecular docking at the in-silico level. It is crucial for studies to report results at least one of these levels. However,

it should be noted that for a drug to be approved for new applications in patients, it must undergo all necessary levels of evaluation.

6 Conclusion

Drug repurposing refers to the reutilization of an existing compound for a therapeutic application other than the one it was originally designed for. This approach has gained increasing importance due to the high costs and lengthy timelines associated with drug development, as well as the need to address orphan diseases and the emergence of new illnesses. In this article, the methods of drug repurposing using graph neural networks have been reviewed. Since this problem is well-suited for modeling with graphs, GNNs are highly appropriate as they combine the advantages of network-based methods with those of the deep learning methods as the most successful artificial intelligence approaches. The article proceeds by presenting a taxonomy to categorize the existing studies. In this taxonomy, GNN-based methods for drug repurposing are classified into eight general categories. Then, each approach and the models developed within its framework were introduced.

Furthermore, the article provides an overview of commonly used datasets, with a comprehensive list included in the appendix. Additionally, the implementation techniques employed in the reviewed studies are summarized, accompanied by a table with links to the corresponding codes of these implementations. The methods used for evaluation of the models are also discussed in detail. The article concludes by outlining a suggested approach for developing GNN-based models for drug repurposing and highlighting key considerations for conducting research in this field.

Appendix 1. Datasets

In this appendix, we introduce several common and useful datasets frequently employed in drug repurposing research. The datasets discussed in previously mentioned papers are included in this list, which represents some of the most commonly used ones. Although the list is extensive, it is not exhaustive of all datasets utilized in drug repurposing studies. The papers [130] and [131] provide particularly informative insights on drug repurposing datasets, and we have relied on these articles to compile this appendix. We strongly recommend reading these papers before selecting any datasets, as they offer valuable guidance.

● **BindingDB**: This is an openly accessible repository that provides experimentally determined binding affinities between proteins and ligands.[132]

● **BioGRID**: A repository focused on compiling molecular interaction data, built through extensive curation efforts.[133]

● **CCLE**: The Cancer Cell Line Encyclopedia (CCLE) offers a diverse collection of cancer cell lines, capturing the genomic variability seen in human cancers. It provides crucial insights into drug responses by including data on genes, mutations, distribution, and methylation scores. [134]

● **CellMinerCDB**: CellMinerCDB is a user-friendly web application designed for easy access and analysis of pharmacogenomic data from cancer cell lines across different platforms. [135]

● **ChEMBL**: ChEMBL is a carefully curated resource containing bioactive molecules with drug-like properties. It integrates details on chemical structures, bioactivity, clinical studies, and disease associations to support drug discovery. [136]

● **ChemDB**: ChemDB is a vast chemical database with around 65 million chemical structures, offering both experimental and predicted molecular properties, including 3D molecular data. [137]

● **ChemicalChecker**: A platform offering harmonized and standardized bioactivity data across different chemical compounds. [138]

● **CTD**: The Comparative Toxicogenomics Database (CTD) organizes and integrates data on chemical interactions with genes proteins, along with chemical-disease and gene-disease relationships. [139]

● **DGIdb**: This dataset compiles drug-target interactions, focusing on the druggable genome, from more than 30 reliable sources like DrugBank, PharmGKB, ChEMBL, and others. [140]

● **DISNET**: The database consists of phenotypes, drugs, pathways, and proteins.[141]

● **DisGeNET**: A platform designed for the discovery and integration of publicly available gene and variant data linked to human diseases. [142]

● **DrugBank**: DrugBank integrates extensive information on drugs, including chemical, pharmacological, and pharmaceutical details, along with the corresponding target information like sequence, structure, and pathways. [143]

● **DrugCentral**: Provides comprehensive data on active chemical entities, their pharmaceutical products, drug mechanisms of action, indications, and pharmacological properties. [144]

● **DTC**: Drug Target Commons (DTC) is a manually curated dataset that includes bioactivity data for proteins, grouped by protein families, clinical phases, side effects, disease targets, and gene associations for approximately 3,000 proteins. [145]

● **GeneCards**: This resource automatically gathers and integrates a wide array of data covering genomics, transcriptomics, proteomics, and clinical information to create a gene-centered resource. [146]

● **GLIDA**: A specialized database that collects drug-target interaction data for G-protein-coupled receptors (GPCRs), focusing specifically on these receptors and their respective ligands. [147]

● **GtopDB**: Provides quantitative bioactivity data for approved and investigational compounds, along with information on drug targets and experimental drugs. [148]

● **Hetnet**: This dataset represents biological data as a knowledge graph, comprising eleven types of nodes and twenty-four types of edges. The node types include anatomy, biological process, cellular component, compound, disease, gene, molecular function, pathway, pharmacologic class, symptom, and side effects.[106]

● **HPRD**: A centralized resource for visualizing and integrating information on post-translational modifications, protein interaction networks, and disease associations for proteins in the human proteome. [149]

● **IntAct**: An open-source platform offering tools and a database for the analysis of molecular interaction data.[150]

● **KEGG**: The Kyoto Encyclopedia of Genes and Genomes (KEGG) serves as a comprehensive knowledge base, integrating genomic, chemical, and functional information for a systematic understanding of gene functions, biological pathways, and their links to diseases and drugs. [151]

● **NCBI GEO**: A repository supporting genomics data, offering access to array- and sequence-based data, allowing users to query and download curated gene expression datasets. [152]

● **OMIM**: OMIM is a widely regarded compendium of human genes and genetic phenotypes, updated daily, with detailed overviews on Mendelian disorders, highlighting the genotype-phenotype relationship. [153]

● **PathBank**: A specialized database that supports pathway analysis and discovery in various fields such as transcriptomics, proteomics, and metabolomics. [154]

● **PathwayCommon**: Stores data on pathways that involve biochemical reactions, complex assembly, and interactions among proteins, DNA, RNA, small molecules, and complexes. [155]

● **PharmGKB**: A comprehensive resource focused on genetic variations and their effects on drug responses, aimed at enhancing clinical practice and research. [156]

● **Probes & Drugs Portal**: A publicly available resource that merges bioactive compound libraries, including screening collections, to provide a comprehensive catalog for drug discovery. [157]

● **PubChem**: An extensive repository of molecular data, including chemical structures, properties, biological activities, safety data, toxicity information, and patent details. [158]

● **RepoDB**: Contains a standardized set of drug repurposing success and failure cases, providing a reliable benchmark for computational drug repurposing studies. [159]

● **STITCH**: A database that provides both known and predicted interactions between chemicals and proteins, including data on pathways and drug-drug interactions. [160]

● **STRING**: Collects and integrates publicly available data on protein-protein interactions and assigns confidence scores to these interactions. [161]

● **Supertarget**: Facilitates the exploration of drug-target interactions and the analysis of associated side effects. [162]

● **SwissTarget Prediction**: Offers information on predicted drug targets based on reverse screening through structural similarity principles.

● **TTD**: The Therapeutic Target Database (TTD) is a repository offering comprehensive details on therapeutic protein and nucleic acid targets, their associated diseases, pathways, and the drugs designed for those targets.

● **UniProt Database**: A thorough resource providing data on proteins, including gene ontology, pathways, taxonomy, and disease-related information.

Appendix 2. Implementation details

This appendix presents a table containing the code repositories, programming languages, and libraries employed for articles with official implementations. The articles are categorized according to the proposed taxonomy.

Table 1. Information about implementations of different GNN approaches to drug repurposing problems

Taxonomy	model name	paper title	link of code	Python					MATLAB
				Torch	PyG	DGL	DeepSnap	TensorFlow	
Attention	DRGBCN	Enhancing Drug repositioning through local Interactive learning with bilinear attention networks	github.com/huangyueshiqi/DRGBCN	*	*				
Attention	MGATRx	MGATRx: Discovering Drug Repositioning Candidates Using Multi-view Graph Attention	github.com/yellajaswanth/MGATRx	*	*				
Attention	DRAW	DRAW: Computational Drug Repositioning with Attention Walking	ads.yonsei.ac.kr/DRAW/	*	*				
Attention	WIGRL	A weighted integration method based on graph representation learning for drug repositioning	github.com/YuBinLab-QUST/WIGRL	*					
Attention	HGTDR	HGTDR: Advancing drug repurposing with heterogeneous graph transformers	github.com/bcb-sut/HGTDR	*	*				
Disentangled representation learning		Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence	github.com/yejinjkim/drug-repurposing-graph	*	*				

Disentangled representation learning	RAFGAE	Drug repositioning based on residual attention network and free multiscale adversarial training	github.com/ghli16/RAFGAE	*	*				
Disentangled representation learning	DRAGON	DRAGON: Drug Repurposing via Graph Neural Networks with Drug and Protein Embeddings as Features,	medal.ctb.upm.es/internal/ gitlab/disnet/gnns/dragon-gnn	*					
Multimodal and data fusion	DRHGCN	Drug repositioning based on the heterogeneous information fusion graph convolutional network	github.com/TheWall9/DRHGCN	*	*				
Multimodal and data fusion	BiFusion	Toward heterogeneous information fusion: bipartite graph convolutional networks for in silico drug repurposing	github.com/zcwang0702/BiFusion	*	*				
Multimodal and data fusion	GCMM	GCMM: graph convolution network based on multimodal attention mechanism for drug repurposing	github.com/FanZhang0820/GCMM						
Multimodal and data fusion	MilGNet	MilGNet: A Multi-instance Learning-based Heterogeneous Graph Network for Drug repositioning	github.com/gu-yaowen/MilGNet	*		*			
Multimodal and data fusion	STRGNN	Deep learning of multimodal networks with topological regularization for drug repositioning	github.com/yuto-ohnuki/STRGNN	*	*				

Multimodal and data fusion	DRMAHGC	Drug repositioning by multi-aspect heterogeneous graph contrastive learning and positive-fusion negative sampling strategy	github.com/JK-Liu7/DRMAHGC	*					
Structural	HGNN-DR	Drug repurposing for Leishmaniasis with Hyperbolic Graph Neural Networks	github.com/layer6ai-labs/hgnn-dr	*					
Structural	WHAIMC	Weighted hypergraph learning and adaptive inductive matrix completion for SARS-CoV-2 drug repositioning	github.com/Mayingjun20179/WHAIMC						*
Structural	PSGCN	PSGCN: Partner-specific drug repositioning approach based on graph convolutional networ	github.com/xinliangSun/PSGCN	*	*				
Structural	CBR-SUBG	Explaining Drug Repositioning: A Case-Based Reasoning Graph Neural Network Approach	github.com/Carolina1396/CBR-SUBG_MIND	*	*				
Data efficient	SADR	SADR: self-supervised graph learning with adaptive denoising for drug repositioning	github.com/Soar1998/SADR	*					
Data efficient	DRGCL	Semantic-enhanced Graph Contrastive Learning with Adaptive Denoising for Drug Repositioning	github.com/Jiaxiao123/DRGCL	*		*			

Data efficient	SGCD	Semantic-enhanced Graph Contrastive Learning with Adaptive Denoising for Drug Repositioning	github.com/yuhuimin11/SGCD-master	*					
Data efficient	TXGNN	Zero-shot drug repurposing with geometric deep learning and clinician centered design	github.com/mims-harvard/TxGNN	*		*			
Explainable AI	GraphIX	GraphIX: Graph-based In silico XAI(explainable artificial intelligence) for drug repositioning from biopharmaceutical network	github.com/TeamGraphix/graphix					*	
Explainable AI	iDPath	Deep learning identifies explainable reasoning paths of mechanism of action for drug repurposing from multilayer biological network	github.com/JasonJYang/iDPath	*					
Explainable AI	XG4Repo	Explainable drug repurposing via path based knowledge graph completion	github.com/AnaJimBej/XG4Repo	*					
RL	KGML-xDTD	KGML-xDTD: a knowledge graph–based machine learning framework for drug treatment prediction and mechanism description	github.com/chunyuma/KGML-xDTD	*					

LLM	Lovenet	All You Need is LOVE: Large Optimized Vector Embeddings Network for Drug Repurposing	github.com/KlickInc/brave-foundry-drug-repurposing	*					
LLM	LLM-DDA	Empowering Graph Neural Network-Based Computational Drug Repositioning with Large Language Model-Inferred Knowledge Representation	github.com/Somewhat120/LLM_DDA	*		*			

References

[1] Hui-Hsing Wong and Amber Jessup, "Examination of Clinical Trial Costs and Barriers for Drug Development." [Online]. Available: https://aspe.hhs.gov/reports/examination-clinical-trial-costs-barriers-drug-development-0

[2] Y. Yeu, Y. Yoon, and S. Park, "Protein localization vector propagation: a method for improving the accuracy of drug repositioning," *Mol. BioSyst.*, vol. 11, no. 7, pp. 2096–2102, 2015, doi: 10.1039/C5MB00306G.

[3] P. P. Deotarse, A. S. Jain, M. B. Baile, N. S. Kolhe, and A. A. Kulkarni, "Drug Repositioning: A Review," *International Journal of Pharma Research & Review, August 2015*, 2015.

[4] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nat Rev Drug Discov*, vol. 3, no. 8, pp. 673–683, Aug. 2004, doi: 10.1038/nrd1468.

[5] B. Shaker, S. Ahmad, J. Lee, C. Jung, and D. Na, "In silico methods and tools for drug discovery," *Computers in Biology and Medicine*, vol. 137, p. 104851, Oct. 2021, doi: 10.1016/j.compbiomed.2021.104851.

[6] B. S. Robinson *et al.*, "BMS-232632, a Highly Potent Human Immunodeficiency Virus Protease Inhibitor That Can Be Used in Combination with Other Available Antiretroviral Agents," *Antimicrob Agents Chemother*, vol. 44, no. 8, pp. 2093–2099, Aug. 2000, doi: 10.1128/AAC.44.8.2093-2099.2000.

[7] A. Krohn, S. Redshaw, J. C. Ritchie, B. J. Graves, and M. H. Hatada, "Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere," *J. Med. Chem.*, vol. 34, no. 11, pp. 3340–3342, Nov. 1991, doi: 10.1021/jm00115a028.

[8] H. Jhoti and A. R. Leach, *Structure-Based Drug Discovery*. Dordrecht: Springer Netherlands, 2007. doi: 10.1007/1-4020-4407-0.

[9] D. Vidal, R. Garcia-Serna, and J. Mestres, "Ligand-Based Approaches to In Silico Pharmacology," in *Chemoinformatics and Computational Chemical Biology*, vol. 672, J. Bajorath, Ed., in Methods in Molecular Biology, vol. 672. , Totowa, NJ: Humana Press, 2011, pp. 489–502. doi: 10.1007/978-1-60761-839-3_19.

[10] Z. Liu *et al.*, "In silico drug repositioning – what we need to know," *Drug Discovery Today*, vol. 18, no. 3–4, pp. 110–115, Feb. 2013, doi: 10.1016/j.drudis.2012.08.005.

[11] J. Aldana-Bitar *et al.*, "Colchicine and plaque: A focus on atherosclerosis imaging," *Progress in Cardiovascular Diseases*, p. S0033062024000318, Feb. 2024, doi: 10.1016/j.pcad.2024.02.010.

[12] Y. Ge *et al.*, "An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19," *Sig Transduct Target Ther*, vol. 6, no. 1, p. 165, Apr. 2021, doi: 10.1038/s41392-021-00568-6.

[13] D. Scherman and C. Fetro, "Drug repositioning for rare diseases: Knowledge-based success stories," *Therapies*, vol. 75, no. 2, pp. 161–167, Apr. 2020, doi: 10.1016/j.therap.2020.02.007.

[14] A. Barbarossa *et al.*, "Non-Antibiotic Drug Repositioning as an Alternative Antimicrobial Approach," *Antibiotics*, vol. 11, no. 6, p. 816, Jun. 2022, doi: 10.3390/antibiotics11060816.

[15] Y. Lau, J. M. Gutierrez, M. Volkovs, and S. Zuberi, "Drug repurposing for Leishmaniasis with Hyperbolic Graph Neural Networks," Feb. 13, 2023. doi: 10.1101/2023.02.11.528117.

[16] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, and J. Wang, "Computational drug repositioning using low-rank matrix approximation and randomized algorithms," *Bioinformatics*, vol. 34, no. 11, pp. 1904–1912, Jun. 2018, doi: 10.1093/bioinformatics/bty013.

[17] A. Zabihian, J. Asghari, M. Hooshmand, and S. Gharaghani, "A comparative analysis of computational drug repurposing approaches: proposing a novel tensor-matrix-tensor factorization method," *Mol Divers*, vol. 28, no. 4, pp. 2177–2196, Aug. 2024, doi: 10.1007/s11030-024-10851-7.

[18] X. Wang and R. Yan, "DDAPRED: a computational method for predicting drug repositioning using regularized logistic matrix factorization," *J Mol Model*, vol. 26, no. 3, p. 60, Mar. 2020, doi: 10.1007/s00894-020-4315-x.

[19] Y. Meng, M. Jin, X. Tang, and J. Xu, "Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study," *Applied Soft Computing*, vol. 103, p. 107135, May 2021, doi: 10.1016/j.asoc.2021.107135.

[20] J. Yang, Z. Li, X. Fan, and Y. Cheng, "Drug–Disease Association and Drug-Repositioning Predictions in Complex Diseases Using Causal Inference–Probabilistic Matrix Factorization," *J. Chem. Inf. Model.*, vol. 54, no. 9, pp. 2562–2569, Sep. 2014, doi: 10.1021/ci500340n.

[21] A.-L. Barabási, *Linked: the new science of networks*. Cambridge, Mass: Perseus Pub, 2002.

[22] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999, doi: 10.1126/science.286.5439.509.

[23] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998, doi: 10.1038/30918.

[24] M. Mitchell, *Complexity: a guided tour*, 1. iss. as an paperback. New York, NY: Oxford University Press, 2011.

[25] V. Martínez, C. Navarro, C. Cano, W. Fajardo, and A. Blanco, "DrugNet: Network-based drug–disease prioritization by integrating heterogeneous data," *Artificial Intelligence in Medicine*, vol. 63, no. 1, pp. 41–49, Jan. 2015, doi: 10.1016/j.artmed.2014.11.003.

[26] M. Lotfi Shahreza, N. Ghadiri, S. R. Mousavi, J. Varshosaz, and J. R. Green, "A review of network-based approaches to drug repositioning," *Briefings in Bioinformatics*, vol. 19, no. 5, pp. 878–892, Sep. 2018, doi: 10.1093/bib/bbx017.

[27] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*, Fourth Edition. in Pearson Series in Artificial Intelligence. Hoboken, NJ: Pearson, 2021.

[28] C. M. Bishop, *Pattern recognition and machine learning*. in Information Science and Statistics. New York, NY: Springer Science+Business Media, LLC, 2019.

[29] S. Arlot and R. Genuer, "Comments on: A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 228–238, Jun. 2016, doi: 10.1007/s11749-016-0484-4.

[30] A. Gottlieb, G. Y. Stein, E. Ruppin, and R. Sharan, "PREDICT: a method for inferring novel drug indications with application to personalized medicine," *Molecular Systems Biology*, vol. 7, no. 1, p. 496, Jan. 2011, doi: 10.1038/msb.2011.26.

[31] Y. Wang, S. Chen, N. Deng, and Y. Wang, "Drug Repositioning by Kernel-Based Integration of Molecular Structure, Molecular Activity, and Phenotype Data," *PLoS ONE*, vol. 8, no. 11, p. e78518, Nov. 2013, doi: 10.1371/journal.pone.0078518.

[32] D. Cao *et al.*, "Computational Prediction of Drug□Target Interactions Using Chemical, Biological, and Network Features," *Molecular Informatics*, vol. 33, no. 10, pp. 669–681, Oct. 2014, doi: 10.1002/minf.201400009.

[33] C. M. Bishop and H. Bishop, *Deep Learning: Foundations and Concepts*. Cham: Springer International Publishing, 2024. doi: 10.1007/978-3-031-45468-4.

[34] S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An Attentive Survey of Attention Models," Jul. 12, 2021, *arXiv*: arXiv:1904.02874. Accessed: Jun. 11, 2024. [Online]. Available: http://arxiv.org/abs/1904.02874

[35] L. Wu, P. Cui, J. Pei, and L. Zhao, Eds., *Graph neural networks: foundations, frontiers, and applications*. Singapore: Springer, 2022. doi: 10.1007/978-981-16-6054-2.

[36] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," 2017, *arXiv*. doi: 10.48550/ARXIV.1710.10903.

[37] A. Vaswani *et al.*, "Attention Is All You Need," 2017, *arXiv*. doi: 10.48550/ARXIV.1706.03762.

[38] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim, "Graph Transformer Networks," 2019, *arXiv*. doi: 10.48550/ARXIV.1911.06455.

[39] X. Tang *et al.*, "Enhancing Drug Repositioning through Local Interactive Learning with Bilinear Attention Networks," *IEEE J. Biomed. Health Inform.*, pp. 1–12, 2024, doi: 10.1109/JBHI.2023.3335275.

[40] J. K. Yella and A. G. Jegga, "MGATRx: Discovering Drug Repositioning Candidates Using Multi-View Graph Attention," *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 19, no. 5, pp. 2596–2604, Sep. 2022, doi: 10.1109/TCBB.2021.3082466.

[41] X. Wang *et al.*, "Drug repositioning of COVID-19 based on mixed graph network and ion channel," *MBE*, vol. 19, no. 4, pp. 3269–3284, 2022, doi: 10.3934/mbe.2022151.

[42] "rdki." [Online]. Available: https://www.rdkit.org/

[43] X. Tian and Y. Meng, "Relgraph: A Multi-Relational Graph Neural Network Framework for Knowledge Graph Reasoning Based on Relation Graph," *Applied Sciences*, vol. 14, no. 7, p. 3122, Apr. 2024, doi: 10.3390/app14073122.

[44] T. Shen, F. Zhang, and J. Cheng, "A comprehensive overview of knowledge graph completion," *Knowledge-Based Systems*, vol. 255, p. 109597, Nov. 2022, doi: 10.1016/j.knosys.2022.109597.

[45] J.-H. Park and Y.-R. Cho, "Computational drug repositioning with attention walking," *Sci Rep*, vol. 14, no. 1, p. 10072, May 2024, doi: 10.1038/s41598-024-60756-6.

[46] J. Tang, X. Zeng, P. Li, W. Chen, C.-M. Liu, and B. Lei, "GTDDA: Graph Convolutional Network and Graph Transformer Structure for Drug Repositioning," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Istanbul, Turkiye: IEEE, Dec. 2023, pp. 436–439. doi: 10.1109/BIBM58861.2023.10386002.

[47] H. Lian, P. Ding, C. Yu, X. Zhang, G. Liu, and B. Yu, "A weighted integration method based on graph representation learning for drug repositioning," *Applied Soft Computing*, vol. 161, p. 111763, Aug. 2024, doi: 10.1016/j.asoc.2024.111763.

[48] H. He *et al.*, "DRTerHGAT: A drug repurposing method based on the ternary heterogeneous graph attention network," *Journal of Molecular Graphics and Modelling*, vol. 130, p. 108783, Jul. 2024, doi: 10.1016/j.jmgm.2024.108783.

[49] A. Gharizadeh, K. Abbasi, A. Ghareyazi, M. R. K. Mofrad, and H. R. Rabiee, "HGTDR: Advancing drug repurposing with heterogeneous graph transformers," *Bioinformatics*, vol. 40, no. 7, p. btae349, Jul. 2024, doi: 10.1093/bioinformatics/btae349.

[50] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous Graph Transformer," Mar. 02, 2020, *arXiv*: arXiv:2003.01332. Accessed: Aug. 12, 2024. [Online]. Available: http://arxiv.org/abs/2003.01332

[51] P. Chandak, K. Huang, and M. Zitnik, "Building a knowledge graph to enable precision medicine," *Sci Data*, vol. 10, no. 1, p. 67, Feb. 2023, doi: 10.1038/s41597-023-01960-3.

[52] S. Chithrananda, G. Grand, and B. Ramsundar, "ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction," Oct. 23, 2020, *arXiv*: arXiv:2010.09885. Accessed: Aug. 12, 2024. [Online]. Available: http://arxiv.org/abs/2010.09885

[53] J. Lee *et al.*, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 2020, doi: 10.1093/bioinformatics/btz682.

[54] X. Wang, H. Chen, S. Tang, Z. Wu, and W. Zhu, "Disentangled Representation Learning," 2022, *arXiv*. doi: 10.48550/ARXIV.2211.11695.

[55] R. Salakhutdinov and G. Hinton, "Semantic hashing," *International Journal of Approximate Reasoning*, vol. 50, no. 7, pp. 969–978, Jul. 2009, doi: 10.1016/j.ijar.2008.11.006.

[56] D. Wang, P. Cui, and W. Zhu, "Structural Deep Network Embedding," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 1225–1234. doi: 10.1145/2939672.2939753.

[57] D. P. Kingma and M. Welling, "An Introduction to Variational Autoencoders," *FNT in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019, doi: 10.1561/2200000056.

[58] T. N. Kipf and M. Welling, "Variational Graph Auto-Encoders," 2016, *arXiv*. doi: 10.48550/ARXIV.1611.07308.

[59] S. Lei, X. Lei, and L. Liu, "Drug repositioning based on heterogeneous networks and variational graph autoencoders," *Front. Pharmacol.*, vol. 13, p. 1056605, Dec. 2022, doi: 10.3389/fphar.2022.1056605.

[60] K. Hsieh *et al.*, "Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence," *Sci Rep*, vol. 11, no. 1, p. 23179, Nov. 2021, doi: 10.1038/s41598-021-02353-5.

[61] Y. Zhang, X. Lei, Y. Pan, and F.-X. Wu, "Drug Repositioning with GraphSAGE and Clustering Constraints Based on Drug and Disease Networks," *Front. Pharmacol.*, vol. 13, p. 872785, May 2022, doi: 10.3389/fphar.2022.872785.

[62] D.-X. Li *et al.*, "A Novel Graph Representation Learning Model for Drug Repositioning Using Graph Transition Probability Matrix Over Heterogenous Information Networks," in *Advanced Intelligent Computing Technology and Applications*, vol. 14088, D.-S. Huang, P. Premaratne, B. Jin, B. Qu, K.-H. Jo, and A. Hussain, Eds., in Lecture Notes in Computer Science, vol. 14088. , Singapore: Springer Nature Singapore, 2023, pp. 180–191. doi: 10.1007/978-981-99-4749-2_16.

[63] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," Jan. 20, 2019, *arXiv*: arXiv:1706.09516. Accessed: Jun. 11, 2024. [Online]. Available: http://arxiv.org/abs/1706.09516

[64] G. Li, S. Li, C. Liang, Q. Xiao, and J. Luo, "Drug repositioning based on residual attention network and free multiscale adversarial training," *BMC Bioinformatics*, vol. 25, no. 1, p. 261, Aug. 2024, doi: 10.1186/s12859-024-05893-5.

[65] R. Artiñano-Muñoz, L. Prieto-Santamaría, A. Pérez-Pérez, and A. Rodríguez-González, "DRAGON: Drug Repurposing via Graph Neural Networks with Drug and Protein Embeddings as Features," in *2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS)*, Guadalajara, Mexico: IEEE, Jun. 2024, pp. 170–175. doi: 10.1109/CBMS61543.2024.00036.

[66] A. Ayuso-Muñoz, L. Prieto-Santamaría, A. Álverez-Pérez, B. Otero-Carrasco, E. Serrano, and A. Rodríguez-González, "Enhancing Drug Repurposing on Graphs by Integrating Drug Molecular Structure as Feature," in *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, L'Aquila, Italy: IEEE, Jun. 2023, pp. 192–197. doi: 10.1109/CBMS58004.2023.00215.

[67] N. García Sánchez, E. Ugarte Carro, L. Prieto-Santamaría, and A. Rodríguez-González, "Protein sequence analysis in the context of drug repurposing," *BMC Med Inform Decis Mak*, vol. 24, no. 1, p. 122, May 2024, doi: 10.1186/s12911-024-02531-1.

[68] G. M. Dimitri, "A Short Survey on Deep Learning for Multimodal Integration: Applications, Future Perspectives and Challenges," *Computers*, vol. 11, no. 11, p. 163, Nov. 2022, doi: 10.3390/computers11110163.

[69] L. Cai *et al.*, "Drug repositioning based on the heterogeneous information fusion graph convolutional network," *Briefings in Bioinformatics*, vol. 22, no. 6, p. bbab319, Nov. 2021, doi: 10.1093/bib/bbab319.

[70] Z. Wang, M. Zhou, and C. Arnold, "Toward heterogeneous information fusion: bipartite graph convolutional networks for *in silico* drug repurposing," *Bioinformatics*, vol. 36, no. Supplement_1, pp. i525–i533, Jul. 2020, doi: 10.1093/bioinformatics/btaa437.

[71] Z. Xiong, F. Huang, Z. Wang, S. Liu, and W. Zhang, "A Multimodal Framework for Improving *in Silico* Drug Repositioning With the Prior Knowledge From Knowledge Graphs," *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 19, no. 5, pp. 2623–2631, Sep. 2022, doi: 10.1109/TCBB.2021.3103595.

[72] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge Graph Embedding via Dynamic Mapping Matrix," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, China: Association for Computational Linguistics, 2015, pp. 687–696. doi: 10.3115/v1/P15-1067.

[73] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *J. Chem. Inf. Comput. Sci.*, vol. 28, no. 1, pp. 31–36, Feb. 1988, doi: 10.1021/ci00057a005.

[74] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation," Jul. 07, 2020, *arXiv*: arXiv:2002.02126. Accessed: Jun. 11, 2024. [Online]. Available: http://arxiv.org/abs/2002.02126

[75] F. Zhang, W. Hu, and Y. Liu, "GCMM: graph convolution network based on multimodal attention mechanism for drug repurposing," *BMC Bioinformatics*, vol. 23, no. 1, p. 372, Sep. 2022, doi: 10.1186/s12859-022-04911-8.

[76] B.-W. Zhao *et al.*, "Fusing Higher and Lower-Order Biological Information for Drug Repositioning via Graph Representation Learning," *IEEE Trans. Emerg. Topics Comput.*, vol. 12, no. 1, pp. 163–176, Jan. 2024, doi: 10.1109/TETC.2023.3239949.

[77] Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: Scalable Representation Learning for Heterogeneous Networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax NS Canada: ACM, Aug. 2017, pp. 135–144. doi: 10.1145/3097983.3098036.

[78] L. Zhang and P. N. Suganthan, "A comprehensive evaluation of random vector functional link networks," *Information Sciences*, vol. 367–368, pp. 1094–1105, Nov. 2016, doi: 10.1016/j.ins.2015.09.025.

[79] Y. Gu, S. Zheng, B. Zhang, H. Kang, and J. Li, "MilGNet: A Multi-instance Learning-based Heterogeneous Graph Network for Drug repositioning," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Las Vegas, NV, USA: IEEE, Dec. 2022, pp. 430–437. doi: 10.1109/BIBM55620.2022.9995152.

[80] X. Gao, X. Jia, Y. Li, and H. Xiong, "Dynamic Scenario Representation Learning for Motion Forecasting with Heterogeneous Graph Convolutional Recurrent Networks," Mar. 07, 2023, *arXiv*: arXiv:2303.04364. Accessed: Jul. 31, 2024. [Online]. Available: http://arxiv.org/abs/2303.04364

[81] Y. Ohnuki, M. Akiyama, and Y. Sakakibara, "Deep learning of multimodal networks with topological regularization for drug repositioning," Jan. 10, 2024. doi: 10.26434/chemrxiv-2024-pd20z.

[82] J. Liu, F. Hu, Q. Zou, P. Tiwari, H. Wu, and Y. Ding, "Drug repositioning by multi-aspect heterogeneous graph contrastive learning and positive-fusion negative sampling strategy," *Information Fusion*, vol. 112, p. 102563, Dec. 2024, doi: 10.1016/j.inffus.2024.102563.

[83] P. Zeng *et al.*, "Drug repositioning based on tripartite cross-network embedding and graph convolutional network," *Expert Systems with Applications*, vol. 252, p. 124152, Oct. 2024, doi: 10.1016/j.eswa.2024.124152.

[84] Q. Liu, M. Nickel, and D. Kiela, "Hyperbolic Graph Neural Networks," 2019, *arXiv*. doi: 10.48550/ARXIV.1910.12892.

[85] I. Chami, R. Ying, C. Ré, and J. Leskovec, "Hyperbolic Graph Convolutional Neural Networks," 2019, *arXiv*. doi: 10.48550/ARXIV.1910.12933.

[86] A. Bretto, *Hypergraph theory: an introduction*. in Mathematical engineering. Cham New York: Springer, 2013.

[87] A. Jain, M. Laure-Charpignon, I. Y. Chen, A. Philippakis, and A. Alaa, "Generating Drug Repurposing Hypotheses through the Combination of Disease-Specific Hypergraphs," 2023, *arXiv*. doi: 10.48550/ARXIV.2311.09596.

[88] A. Grover and J. Leskovec, "node2vec: Scalable Feature Learning for Networks," Jul. 03, 2016, *arXiv*: arXiv:1607.00653. Accessed: Jun. 10, 2024. [Online]. Available: http://arxiv.org/abs/1607.00653

[89] S. Pang *et al.*, "HGDD: A Drug-Disease High-Order Association Information Extraction Method for Drug Repurposing via Hypergraph," in *Bioinformatics Research and Applications*, vol. 13064, Y. Wei, M. Li, P. Skums, and Z. Cai, Eds., in Lecture Notes in Computer Science, vol. 13064. , Cham: Springer International Publishing, 2021, pp. 424–435. doi: 10.1007/978-3-030-91415-8_36.

[90] W. Huang, Z. Li, Y. Kang, X. Ye, and W. Feng, "Drug Repositioning Based on the Enhanced Message Passing and Hypergraph Convolutional Networks," *Biomolecules*, vol. 12, no. 11, p. 1666, Nov. 2022, doi: 10.3390/biom12111666.

[91] X. Sun, B. Wang, J. Zhang, and M. Li, "Partner-Specific Drug Repositioning Approach Based on Graph Convolutional Network," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 11, pp. 5757–5765, Nov. 2022, doi: 10.1109/JBHI.2022.3194891.

[92] Carolina González-Cavazos, "Explaining Drug Repositioning: A Case-Based Reasoning Graph Neural Network Approach".

[93] R. Balestriero *et al.*, "A Cookbook of Self-Supervised Learning," 2023, *arXiv*. doi: 10.48550/ARXIV.2304.12210.

[94] S. Jin, Y. Zhang, H. Yu, and M. Lu, "SADR: Self-Supervised Graph Learning With Adaptive Denoising for Drug Repositioning," *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 21, no. 2, pp. 265–277, Mar. 2024, doi: 10.1109/TCBB.2024.3351079.

[95] X. Jia, X. Sun, K. Wang, and M. Li, "DRGCL: Drug Repositioning via Semantic-enriched Graph Contrastive Learning," *IEEE J. Biomed. Health Inform.*, pp. 1–12, 2024, doi: 10.1109/JBHI.2024.3372527.

[96] H. Yu, M. Lu, Z. Li, and Y. Zhang, "Semantic-enhanced Graph Contrastive Learning with Adaptive Denoising for Drug Repositioning," *IEEE J. Biomed. Health Inform.*, pp. 1–9, 2024, doi: 10.1109/JBHI.2023.3344031.

[97] Y. Song, T. Wang, S. K. Mondal, and J. P. Sahoo, "A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities," 2022, *arXiv*. doi: 10.48550/ARXIV.2205.06743.

[98] C. Yang, X. Chen, J. Huang, Y. An, Z. Huang, and Y. Sun, "A few-shot link prediction framework to drug repurposing using multi-level attention network," *Computers in Biology and Medicine*, vol. 170, p. 107936, Mar. 2024, doi: 10.1016/j.compbiomed.2024.107936.

[99] V. N. Ioannidis, D. Zheng, and G. Karypis, "Few-shot link prediction via graph neural networks for Covid-19 drug-repurposing," 2020, *arXiv*. doi: 10.48550/ARXIV.2007.10261.

[100] K. Huang *et al.*, "Zero-shot drug repurposing with geometric deep learning and clinician centered design," Apr. 29, 2024, *medRxiv*. doi: 10.1101/2023.03.19.23287458.

[101] S. R. Islam, W. Eberle, S. K. Ghafoor, and M. Ahmed, "Explainable Artificial Intelligence Approaches: A Survey," Jan. 23, 2021, *arXiv*: arXiv:2101.09429. Accessed: Jun. 11, 2024. [Online]. Available: http://arxiv.org/abs/2101.09429

[102] A. Takagi, M. Kamada, E. Hamatani, R. Kojima, and Y. Okuno, "GraphIX: Graph-based In silico XAI(explainable artificial intelligence) for drug repositioning from biopharmaceutical network," Dec. 27, 2022, *arXiv*: arXiv:2212.10788. Accessed: Jun. 11, 2024. [Online]. Available: http://arxiv.org/abs/2212.10788

[103] C. He, L. Duan, H. Zheng, L. Song, and M. Huang, "An explainable framework for drug repositioning from disease information network," *Neurocomputing*, vol. 511, pp. 247–258, Oct. 2022, doi: 10.1016/j.neucom.2022.09.063.

[104] J. Yang *et al.*, "Deep learning identifies explainable reasoning paths of mechanism of action for drug repurposing from multilayer biological network," *Briefings in Bioinformatics*, vol. 23, no. 6, p. bbac469, Nov. 2022, doi: 10.1093/bib/bbac469.

[105] A. Jiménez, M. J. Merino, J. Parras, and S. Zazo, "Explainable drug repurposing via path based knowledge graph completion," *Sci Rep*, vol. 14, no. 1, p. 16587, Jul. 2024, doi: 10.1038/s41598-024-67163-x.

[106] D. S. Himmelstein *et al.*, "Hetnet connectivity search provides rapid insights into how biomedical entities are related," *GigaScience*, vol. 12, p. giad047, Dec. 2022, doi: 10.1093/gigascience/giad047.

[107] C. Ma, Z. Zhou, H. Liu, and D. Koslicki, "KGML-xDTD: a knowledge graph–based machine learning framework for drug treatment prediction and mechanism description," *GigaScience*, vol. 12, p. giad057, Dec. 2022, doi: 10.1093/gigascience/giad057.

[108] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," 2017, *arXiv*. doi: 10.48550/ARXIV.1706.02216.

[109] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Oct. 16, 2013, *arXiv*: arXiv:1310.4546. Accessed: Jul. 29, 2024. [Online]. Available: http://arxiv.org/abs/1310.4546

[110] S. Minaee *et al.*, "Large Language Models: A Survey," Feb. 20, 2024, *arXiv*: arXiv:2402.06196. Accessed: Oct. 09, 2024. [Online]. Available: http://arxiv.org/abs/2402.06196

[111] Sepehr Asgarian, Sina Akbarian, and Jouhyun Jeon, "All You Need is LOVE: Large Optimized Vector Embeddings Network for Drug Repurposing".

[112] H. Touvron *et al.*, "Llama 2: Open Foundation and Fine-Tuned Chat Models," Jul. 19, 2023, *arXiv*: arXiv:2307.09288. Accessed: Jun. 11, 2024. [Online]. Available: http://arxiv.org/abs/2307.09288

[113] Y. Gu, Z. Xu, and C. Yang, "Empowering Graph Neural Network-Based Computational Drug Repositioning with Large Language Model-Inferred Knowledge Representation," *Interdiscip Sci Comput Life Sci*, Sep. 2024, doi: 10.1007/s12539-024-00654-7.

[114] "pytorch." [Online]. Available: https://pytorch.org/

[115] "PyG." [Online]. Available: https://pytorch-geometric.readthedocs.io/en/latest/

[116] "tensorflow." [Online]. Available: https://www.tensorflow.org/

[117] "DGL." [Online]. Available: https://www.dgl.ai/

[118] "deepsnap." [Online]. Available: https://snap.stanford.edu/deepsnap/

[119] "Graph neural networks in TensorFlow." [Online]. Available: https://blog.tensorflow.org/2024/02/graph-neural-networks-in-tensorflow.html

[120] "Matlab." [Online]. Available: https://www.mathworks.com/company.html

[121] "Matlab and GNN." [Online]. Available: https://www.mathworks.com/help/deeplearning/ug/node-classification-using-graph-convolutional-network.html

[122] J. Haneczok, M. Delijewski, and R. Moldzio, "AI molecular property prediction for Parkinson's Disease reveals potential repurposing drug candidates based on the increase of the expression of PINK1," *Computer Methods and Programs in Biomedicine*, vol. 241, p. 107731, Nov. 2023, doi: 10.1016/j.cmpb.2023.107731.

[123] L. Yu, J. Zhao, and L. Gao, "Drug repositioning based on triangularly balanced structure for tissue-specific diseases in incomplete interactome," *Artificial Intelligence in Medicine*, vol. 77, pp. 53–63, Mar. 2017, doi: 10.1016/j.artmed.2017.03.009.

[124] C. Bansal, R. Chandra, V. Agarwal, and P. R. Deepa, "A clustering and graph deep learning-based framework for COVID-19 drug repurposing," Jun. 24, 2023, *arXiv*: arXiv:2306.13995. doi: 10.48550/arXiv.2306.13995.

[125] B. G. Katzung and T. W. Vanderah, Eds., *Basic & clinical pharmacology*, Fifteenth edition. in A Lange medical book. New York Chicago San Francisco Athens London Madrid Mexico City Milan New Delhi Singapore Sydney Toronto: McGraw-Hill, 2021.

[126] M. M. Coelho, C. Fernandes, F. Remião, and M. E. Tiritan, "Enantioselectivity in Drug Pharmacokinetics and Toxicity: Pharmacological Relevance and Analytical Methods," *Molecules*, vol. 26, no. 11, p. 3113, May 2021, doi: 10.3390/molecules26113113.

[127] S. Luo *et al.*, "One Transformer Can Understand Both 2D & 3D Molecular Data," Mar. 28, 2023, *arXiv*: arXiv:2210.01765. doi: 10.48550/arXiv.2210.01765.

[128] T.-H. Nguyen-Vo, P. Teesdale-Spittle, J. E. Harvey, and B. P. Nguyen, "Molecular representations in bio-cheminformatics," *Memetic Comp.*, vol. 16, no. 3, pp. 519–536, Sep. 2024, doi: 10.1007/s12293-024-00414-6.

[129] C. A. Gebeshuber *et al.*, "Computational drug repositioning of clopidogrel as a novel therapeutic option for focal segmental glomerulosclerosis," *Translational Research*, vol. 259, pp. 28–34, Sep. 2023, doi: 10.1016/j.trsl.2023.04.001.

[130] Z. Tanoli, U. Seemab, A. Scherer, K. Wennerberg, J. Tang, and M. Vähä-Koskela, "Exploration of databases and methods supporting drug repurposing: a comprehensive survey," *Briefings in Bioinformatics*, vol. 22, no. 2, pp. 1656–1678, Mar. 2021, doi: 10.1093/bib/bbaa003.

[131] X. Pan *et al.*, "Deep learning for drug repurposing: Methods, databases, and applications," *WIREs Comput Mol Sci*, vol. 12, no. 4, p. e1597, Jul. 2022, doi: 10.1002/wcms.1597.

[132] M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, and J. Chong, "BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology," *Nucleic Acids Res*, vol. 44, no. D1, pp. D1045–D1053, Jan. 2016, doi: 10.1093/nar/gkv1072.

[133] R. Oughtred *et al.*, "The BioGRID interaction database: 2019 update," *Nucleic Acids Research*, vol. 47, no. D1, pp. D529–D541, Jan. 2019, doi: 10.1093/nar/gky1079.

[134] J. Barretina *et al.*, "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, Mar. 2012, doi: 10.1038/nature11003.

[135] V. N. Rajapakse *et al.*, "CellMinerCDB for Integrative Cross-Database Genomics and Pharmacogenomics Analyses of Cancer Cell Lines," *iScience*, vol. 10, pp. 247–264, Dec. 2018, doi: 10.1016/j.isci.2018.11.029.

[136] A. Gaulton *et al.*, "The ChEMBL database in 2017," *Nucleic Acids Res*, vol. 45, no. D1, pp. D945–D954, Jan. 2017, doi: 10.1093/nar/gkw1074.

[137] J. H. Chen, E. Linstead, S. J. Swamidass, D. Wang, and P. Baldi, "ChemDB update—full-text search and virtual chemical space," *Bioinformatics*, vol. 23, no. 17, pp. 2348–2351, Sep. 2007, doi: 10.1093/bioinformatics/btm341.

[138] M. Duran-Frigola *et al.*, "Extending the small-molecule similarity principle to all levels of biology with the Chemical Checker," *Nat Biotechnol*, vol. 38, no. 9, pp. 1087–1096, Sep. 2020, doi: 10.1038/s41587-020-0502-7.

[139] A. P. Davis *et al.*, "The Comparative Toxicogenomics Database: update 2019," *Nucleic Acids Research*, vol. 47, no. D1, pp. D948–D954, Jan. 2019, doi: 10.1093/nar/gky868.

[140] A. H. Wagner *et al.*, "DGIdb 2.0: mining clinically relevant drug–gene interactions," *Nucleic Acids Res*, vol. 44, no. D1, pp. D1036–D1044, Jan. 2016, doi: 10.1093/nar/gkv1165.

[141] G. Lagunes-García, A. Rodríguez-González, L. Prieto-Santamaría, E. P. García Del Valle, M. Zanin, and E. Menasalvas-Ruiz, "DISNET: a framework for extracting phenotypic disease information from public sources," *PeerJ*, vol. 8, p. e8580, Feb. 2020, doi: 10.7717/peerj.8580.

[142] J. Piñero *et al.*, "DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Res*, vol. 45, no. D1, pp. D833–D839, Jan. 2017, doi: 10.1093/nar/gkw943.

[143] D. S. Wishart, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic Acids Research*, vol. 34, no. 90001, pp. D668–D672, Jan. 2006, doi: 10.1093/nar/gkj067.

[144] O. Ursu *et al.*, "DrugCentral 2018: an update," *Nucleic Acids Research*, vol. 47, no. D1, pp. D963–D970, Jan. 2019, doi: 10.1093/nar/gky963.

[145] Z. Tanoli *et al.*, "Drug Target Commons 2.0: a community platform for systematic analysis of drug–target interaction profiles," *Database*, vol. 2018, Jan. 2018, doi: 10.1093/database/bay083.

[146] G. Stelzer *et al.*, "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses," *CP in Bioinformatics*, vol. 54, no. 1, Jun. 2016, doi: 10.1002/cpbi.5.
[147] Y. Okuno *et al.*, "GLIDA: GPCR ligand database for chemical genomics drug discovery database and tools update," *Nucleic Acids Research*, vol. 36, no. Database, pp. D907–D912, Dec. 2007, doi: 10.1093/nar/gkm948.
[148] S. P. Alexander *et al.*, "THE CONCISE GUIDE TO PHARMACOLOGY 2017/18: Overview," *British J Pharmacology*, vol. 174, no. S1, Dec. 2017, doi: 10.1111/bph.13882.
[149] T. S. Keshava Prasad *et al.*, "Human Protein Reference Database--2009 update," *Nucleic Acids Research*, vol. 37, no. Database, pp. D767–D772, Jan. 2009, doi: 10.1093/nar/gkn892.
[150] S. Orchard *et al.*, "The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases," *Nucl. Acids Res.*, vol. 42, no. D1, pp. D358–D363, Jan. 2014, doi: 10.1093/nar/gkt1115.
[151] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, "KEGG: new perspectives on genomes, pathways, diseases and drugs," *Nucleic Acids Res*, vol. 45, no. D1, pp. D353–D361, Jan. 2017, doi: 10.1093/nar/gkw1092.
[152] T. Barrett *et al.*, "NCBI GEO: archive for functional genomics data sets—update," *Nucleic Acids Research*, vol. 41, no. D1, pp. D991–D995, Nov. 2012, doi: 10.1093/nar/gks1193.
[153] J. S. Amberger, C. A. Bocchini, F. Schiettecatte, A. F. Scott, and A. Hamosh, "OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders," *Nucleic Acids Research*, vol. 43, no. D1, pp. D789–D798, Jan. 2015, doi: 10.1093/nar/gku1205.
[154] D. S. Wishart *et al.*, "PathBank: a comprehensive pathway database for model organisms," *Nucleic Acids Research*, vol. 48, no. D1, pp. D470–D478, Jan. 2020, doi: 10.1093/nar/gkz861.
[155] E. G. Cerami *et al.*, "Pathway Commons, a web resource for biological pathway data," *Nucleic Acids Research*, vol. 39, no. Database, pp. D685–D690, Jan. 2011, doi: 10.1093/nar/gkq1039.
[156] M. Hewett, "PharmGKB: the Pharmacogenetics Knowledge Base," *Nucleic Acids Research*, vol. 30, no. 1, pp. 163–165, Jan. 2002, doi: 10.1093/nar/30.1.163.
[157] C. Skuta *et al.*, "Probes & Drugs portal: an interactive, open data resource for chemical biology," *Nat Methods*, vol. 14, no. 8, pp. 759–760, Aug. 2017, doi: 10.1038/nmeth.4365.
[158] Y. Wang *et al.*, "PubChem BioAssay: 2017 update," *Nucleic Acids Res*, vol. 45, no. D1, pp. D955–D963, Jan. 2017, doi: 10.1093/nar/gkw1118.
[159] A. S. Brown and C. J. Patel, "A standard database for drug repositioning," *Sci Data*, vol. 4, no. 1, p. 170029, Mar. 2017, doi: 10.1038/sdata.2017.29.
[160] D. Szklarczyk, A. Santos, C. von Mering, L. J. Jensen, P. Bork, and M. Kuhn, "STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data," *Nucleic Acids Res*, vol. 44, no. D1, pp. D380–D384, Jan. 2016, doi: 10.1093/nar/gkv1277.
[161] D. Szklarczyk *et al.*, "STRING v10: protein–protein interaction networks, integrated over the tree of life," *Nucleic Acids Research*, vol. 43, no. D1, pp. D447–D452, Jan. 2015, doi: 10.1093/nar/gku1003.
[162] N. Hecker *et al.*, "SuperTarget goes quantitative: update on drug-target interactions," *Nucleic Acids Research*, vol. 40, no. D1, pp. D1113–D1117, Jan. 2012, doi: 10.1093/nar/gkr912.