An Analytical Review of Computational Drug Repurposing

Seyedeh Shaghayegh Sadeghi and Mohammad Reza Keyvanpour

Abstract—Drug repurposing is a vital function in pharmaceutical fields and has gained popularity in recent years in both the pharmaceutical industry and research community. It refers to the process of discovering new uses and indications for existing or failed drugs. It is cost-effective and reliable in contrast to experimental drug discovery, which is a costly, time-consuming, and risky process and limited to a relatively small number of targets. Accordingly, a plethora of computational methodologies have been propounded to repurpose drugs on a large scale by utilizing available high throughput data. The available literature, however, lacks a contemporary and comprehensive analysis of the current computational drug repurposing methodologies. In this paper, we presented a systematic analysis of computational drug repurposing which consists of three main sections: Initially, we categorize the computational drug repurposing methods based on their technical approach and artificial intelligence perspective and discuss the strengths and weaknesses of various methods. Secondly, some general criteria are recommended to analyze our proposed categorization. In the third and final section, a qualitative comparison is made between each approach which is a guide to understanding their preference to one another. Further, this systematic analysis can help in the efficient selection and improvement of drug repurposing techniques based on the nature of computational methods implemented on biological resources.

Index Terms—Drug repurposing, machine learning, network analysis. relation extraction

1 Introduction

Since pharmaceutical discovery and development is an extraordinarily complex and costly research encompassing many facets [1], drug repurposing (*introduced by* Ashburn and Thor [2]) has proved to be a preferred alternative strategy to accelerate drug discovery. For instance, it has been estimated that the total average cost of developing a new drug ranges from $0.8 billion to $1.5 billion taking at least 10-17 years to bring a drug to the market [3] (see Fig. 1). On the other hand, drug repurposing (also known as drug repositioning, drug re-profiling, drug redirecting, drug re-tasking, and therapeutic switching) [4] is relatively inexpensive and carries minimal risk due to availability of previous pharmacological, safety, and toxicology data [5], [6]. Further, it has other advantages over de-novo drug development that further enhance our motivation for drug repurposing. For example, another motivation for drug repurposing is that it fits the aims and scopes of personalized and precision medicine [7], [8]. Additionally, it is an excellent opportunity for rare/orphan and neglected diseases with small patient populations in industrialized countries, making it difficult to market drugs that recoup the cost of research and development which are then profitable over the long term [7], [8], [9], [10], [11]. Further, drug repurposing opens up a new avenue for anticancer drug discovery [12], [13], [14] and a beneficial approach to defeat drug resistance, which is one of the main reasons behind reduced drug efficacy, such as the use of non-antibiotic drugs to overcome antimicrobial resistance [17], [18], [19]. In addition to drug repurposing, drug combination predictions are required to identify effective therapies that prevent drug resistance in an effective manner [20], [21]. This interest in early drug discovery stages for using combined drugs or multi-target inhibitors has increasingly evolved, since some diseases are often caused by multiple molecular and environmental interactions. Another reason is that high risk of developing new multi-target drugs [22], [23]. There is a list of successfully repurposed drugs in [5] such as aspirin, where the initial indication of aspirin was "inflammation and pain", yet a new indication of it is "coagulation and stroke". Given the extensive research and the importance of this subject, several authors have recently reviewed different aspects of in silico repurposing approaches [20], [22], [24], [25], [26], [27]. However, current successes in drug repurpsoing have primarily been the result of serendipitous events based on ad hoc clinical observation, unfocused screening, and "happy accidents". Indeed, comprehensive and rational approaches are urgently needed to explore repositioning opportunities [28]. Therefore, a wide range of methods have been proposed for this purpose. Nevertheless, these methods have to overcome some of the challenges in drug repurposing, including technical and non-technical challenges [29].

One of the most important challenges faced by drug repurposing projects is probably related to commercial, regulatory, or intellectual property reasons (Non-Technical challenges) [30]. For instance, there are several strategies which could be used to achieve sufficient IP protection to maximize the market exclusivity of a drug [31]; An example is the combination of two old drugs for a new indication

• *The authors are with the Data Mining Laboratory, Department of Computer Engineering, Alzahra University, Tehran, Iran. E-mail: s.s.sadeghi@student.alzahra.ac.ir, keyvanpour@alzahra.ac.ir.*

Manuscript received 24 Nov. 2018; revised 12 July 2019; accepted 31 July 2019. Date of publication 12 Aug. 2019; date of current version 1 Apr. 2021. (Corresponding author: Mohammad Reza Keyvanpour.)
Digital Object Identifier no. 10.1109/TCBB.2019.2933825

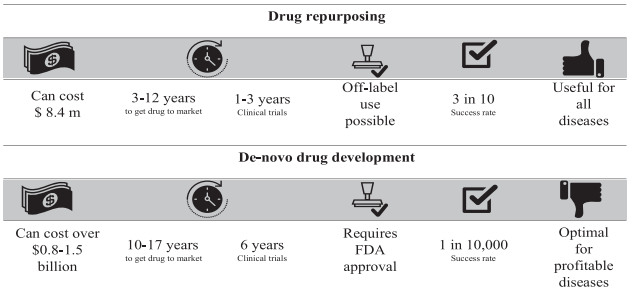

Fig. 1. A short view of the advantages of drug repurposing over De-novo drug development.

[32]. In this regard, a synergistic effect of drug combination therapy can enhance the success rates of drug repurposing and lead to an improved understanding of complicated disease pathophysiology and the design of better treatments for the disease [33]. Meanwhile, the requirement to combine heterogeneous data from multiple sources [26] and selecting of an appropriate method for prediction of new relations; can be considered as a challenging task in computational drug repurposing (Technical challenges).

Considering this problem, we introduced an analytical review in this research, involving three main sections to classify existing computational drug repurposing models. Also, our investigation led to an empirical and technical comparison of computational drug repurposing models. Even prospects and difficulties of each category of models are emphasized for further research in this area. We believe that the three goals above the paper has focused on can be helpful as a practitioner's guide.

The rest of the paper is organized as follows: In the next section, related works are reviewed. In Section 3 we present a formal definition of computational drug repurposing. Section 4 presents the framework of this analytical review. We stress that our categorization of the methods is not clear-cut, as many methods may overlap with each other. We tried to classify them based on what we believe is their main feature/purpose, even if other aspects may be present. Finally, Section 5 deals with the conclusion and future works.

2 Previous Related Review Study

Concerning drug repurposing, a significant body of literature has attempted to propose a framework for drug repurposing. As mentioned earlier, drug repurposing was introduced by Ashburn and Thor [2] and then received attention from other researchers.

An earlier study by Dudley et al. [34] categorized methods of drug repurposing as either 'drug-based', where the discovery of repurposing opportunities initiated from the chemical or pharmaceutical perspective, or 'disease based', where discovery originated from the perspective of disease management, symptomatology or pathology.

Zou et al. [35] proposed another categorization, which categorized drug repurposing methods into 'data-driven', analyzing large-scale '-omics' datasets using statistical modelling techniques, and 'hypothesis-driven' methods, applied to relatively small systems, with often fewer molecular components.

Wu et al. [23] organized network-based methods into two categories, i.e., single drug repositioning and drug combination, and further represented their main features by three data sources. It also discussed the merits and shortcomings of these methods and pinpointed some future topics in this promising field.

Li et al. [22] classified datasets into three strategies: genome, phenome and drug. Then, they categorized drug repurposing methods into 'machine learning', 'network analysis', and 'text mining and semantic inference'. Finally, in addition to discussing validation strategies for repurposing they noted some of the existing challenges in computational drug repurposing.

Vanhaelen et al. [26] proposed a classification of computational drug repurposing methods as '3D structure-based' employing chemical structure files of compounds to compute docking scores; 'similarity-based' utilizing the intuitive notion that similar compounds have similar properties; 'inference-based' benefitting from a network of known interactions to predict new interactions and suggest new targets for drug repositioning; and 'machine learning-based' exploiting similarity measures to construct classification features and subsequent learning of a classification rule which distinguishes true associations of nodes from false ones.

Further, Sam and Athri [36] focused on systematically presenting the available web-based tools that aid in repositioning drugs and provided a classification of web servers to help wet lab researchers choose the right tool to assist studies in repurposing drugs.

Lotfi Shahreza et al. [3] attempted to review network-based methods in predicting drug targets for drug repositioning. For each method, the preferred type of data set was described, and their advantages and limitations were discussed. Also, for each method, a brief description was provided along with an evaluation based on its performance metrics.

Yella et al. [37] presented some of the promising bioinformatics approaches and pipelines, and further summarized and discussed the horizon of computational drug repositioning. Xue et al. [38] reviewed computational approaches and highlighted their characteristics. The computational classification of their strategy was network-based, text mining-based and semantic-based.

In this paper, a comprehensive organization of computational drug repurposing approaches is presented in conjunction with all the previous studies.

3 Drug Repurposing: Problem Definition

Drug repurposing is a strategy to find new uses for previously approved drugs and 'parked' or 'off the shelf' molecules that reached the clinic without any safety concerns but did not show sufficient efficacy against their intended primary disease target [39]. This problem can be formalized as follows:

$$f_{repurposing} : P_{possible} \rightarrow \{0, 1\} \tag{1}$$

$$\begin{aligned} P_{possible} &= D \times R_{FDA\ approved} \\ &= \big\{(d, r) | d \in D \wedge r \in R_{FDA\ approved}\big\}, \end{aligned} \tag{2}$$

Where, $f_{repurposing}$ is the drug repurposing function, with $P_{possible}$ considered denoting the domain of the function and $\{0,\ 1\}$ showing the range of the function. If $p \in P_{possible}$

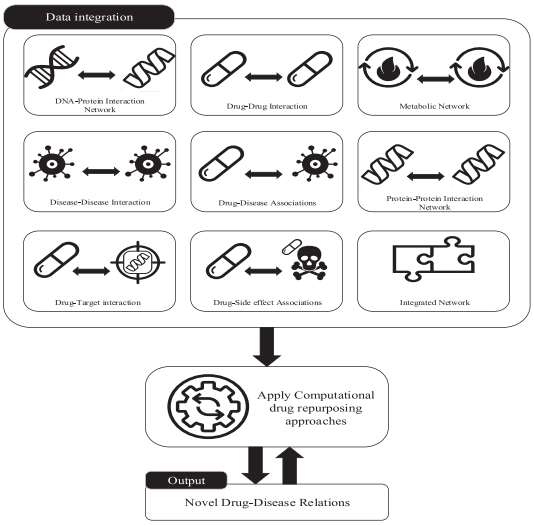

Fig. 2. The general process of computational drug repurposing inspired by [3].

is a known interaction between a disease from the set of D and a drug from the set of $R_{FDA\ approved}$, then the output is 1; otherwise 0. Here, D is a set of all known diseases ranging from any kind of cancer and rare and orphan diseases to very common types of illnesses and $R_{FDA\ approved}$ represents the set of those drugs which are available in the market and shelved.

Drug repurposing can be performed either by an experimental approach, called 'activity-based drug repositioning', or using a computational method called 'in-silico drug repurposing 'or 'computational drug repurposing' [24], [26]. Computational drug repositioning has been shown as a promising and valuable strategy for exploring new uses from existing drugs [40]. To achieve this goal, some computational methods have been proposed, which are based on different types of data sources [40]. Computational drug repurposing approaches differ both in the algorithms they employ as well as in the data sources they utilize [41]. There are many '-omics' molecular data across various levels. The type of data being used and the level that the biological system is under study can dictate which modelling approaches are most appropriate [3]. A comprehensive investigation to gather proper data sources or features related to the problem and to combine them for providing useful information is a decisive step in every computational drug repurposing process. For this reason, integrating information obtained from diverse data sources is essential to enhance information efficiency [42] as each type of data set exposes a unique aspect of information about an organism [43].

It has been shown that drugs with very different chemical structures target the same proteins, where the corresponding protein is druggable from various medications. This fact suggests that drugs are not explicitly designed for diseases [44]; they can also be connected to other networks, associations, and interactions which can affect how we can predict novel drug-disease relations. As illustrated in Fig. 2, other associations can be used alongside the drug and disease information in drug repurposing process to improve the exactitudes of predicted drug-disease relations [3], [30], [45], [46], [47], [48]. Currently, various types of data sources are available as input data for supporting computational drug repositioning. Accordingly, a selection of the frequently used data source types that are useful for computational drug repositioning is summarized in Table 1 [49], [50]. Studies suggest that the integration of drug-target and disease-genes integrations are most popular kind of information integration in drug repurposing methods, such as: TL_HGBI [51], HETER_LP [52], RWHNDR [53], and DrugNet [54]. Nevertheless, choosing the best way for integrating these data sources is a challenging process. Data integration is the most crucial preprocessing step in all biological problems, which is necessary given the large and increasing number of data resources within bioinformatics [55]. Definitely, despite some challenges for data integration, it has various positive points as integrating data from different sources provides us with an unprecedented opportunity to understand a complex biological system from different angles and levels and in turn make precise data-driven predictions [56]. The structure of drug repurposing framework can be presented as a network to predict novel drug-disease relations from integrated data.

A network (or a graph), typically indicated as $G \in (V, E)$, consists of a set of nodes, V, and set of edges, E [92], [93]. A node can represent a biological molecule, e.g., a gene, a protein, a metabolite, or an RNA. Nodes can also be observed at the phenotype level, where they can represent diseases, or drugs. On the other hand, an edge can represent a physical binding (e.g., between a pair of proteins), a functional association (e.g., between a pair of genes), or various types of similarities between nodes (e.g., a chemical similarity between drugs) [94]. Depending on the type of data they represent, network edges can be directed or undirected and weighted or unweighted.

These networks can be shown by a matrix. If nodes are from different types, such as drug-disease relations, this matrix is as below:

$$M_{r_i d_j} = \left\{ \begin{array}{ll} 1 & If\ there\ is\ a\ relation\ between\ r_i\ and\ d_j \\ 0 & Otherwise \end{array} \right\}, \tag{3}$$

Where r_i represents the ith drug and d_j shows the jth disease, with M_{rd} referring to the matrix of the relationship between drugs and diseases. If nodes are of the same types, such as drug-drug relation, each entry of the matrix $M_{r_i d_j} \in [0, 1]$ is a similarity score. These types of structures can be analyzed by network-based methods discussed in the following sections. Meanwhile, the representation of data in drug repurposing systems can be in a text or feature vector format by kernel-based data integration methods. These type of structures can be analyzed by learning-based methods further discussed in the following sections.

In recent years, research on computational drug repurposing has gained remarkable attention and researchers have proposed many methods for application of drug repurposing including the anticancer drug discovery [95], [96], [97] and the discovery of anti-infectious drugs that can overcome drug resistance [98]. Notably, massive feasibility still exists for research in this area.

4 Comparative Analytical Framework for Drug Repurposing

This paper investigates a wide range of computational drug repurposing research to present a comparative analytical

TABLE 1
A Selection of Databases and Resources Available for Computational Drug Repurposing

Type	Resource	Description	URL	Ref
Molecular omics data	**ArrayExpress**	Public repositories of functional genomics data.	https://www.ebi.ac.uk/arrayexpress	[57], [58]
Molecular omics data	**Cancer Cell Line Encyclopedia(CCLE)**	Genomic data of cancer cell lines.	https://portals.broadinstitute.org/ccle	[59]
Chemical structure	**ChEMBL**	Database of over 1 million compound structures, chemical features, bioactivity, etc.	https://www.ebi.ac.uk/chembl	[60], [61]
Adverse effects and clinical trial information	**ClinicalTrials.gov**	A registry and result database of publicly and privately supported clinical studies.	http://www.clinicaltrials.gov	[48]
Chemical–disease–gene association	**Comparative Toxicogenomics Database (CTD)**	It provides curated information about chemical–gene/protein interactions, chemical–disease and gene-disease relationships.	http://ctdbase.org/	[62]
Drug combination–disease association	**Drug Combination Database(DCDB)**	Known examples of drug combinations, models for drug combinations.	http://www.cls.zju.edu.cn/dcdb/	[63]
Signature-matching repositioning pipeline	**Drug versus Disease (DvD)**	Computational pipeline for comparing disease and drug-response gene expression signatures	www.ebi.ac.uk/saezrodriguez/dvd	[64]
Chemical structure and Drug–target information	**DrugBank**	A database that combines detailed drug data with comprehensive drug target information (over 1 million compound structures, chemical features, bioactivity, etc.)	https://www.drugbank.ca/	[65], [66]
Protein interaction information	**STRING**	Include Protein-Protein interaction, analysis, and networks.	https://string-db.org/cgi/input.pl	[67], [68]
Adverse effects and clinical trial information	**FDA Adverse Event Reporting System (FAERS)**	It contains information obtained from adverse event and medication error reports submitted to FDA on side effect Keywords for drugs.	https://open.fda.gov/data/faers/	[69], [70]
Genome information	**Gene Expression Omnibus(GEO)**	An archive of microarray, next-generation sequencing, and other forms of high-throughput functional genomic data submitted by the scientific community, covering a wide variety of experimental conditions including disease characterizations.	http://www.ncbi.nlm.nih.gov/geo	[71], [72]
Pathway information	**Kyoto Encyclopedia of Genes and Genomes (KEGG)**	Resource for understanding high-level functions and utilities of the biological system from molecular-level information.	http://www.genome.jp/kegg/	[73], [74]
Adverse effects and clinical trial information	**Offsides**	Side effects and ADEs not listed on FDA's official drug label.	http://tatonettilab.org/resources/tatonetti-stm.html	[75]
Genome information	**Online Mendelian Inheritance in Man (OMIM)**	Relationships between genes and genetic phenotypes, particularly disorders.	https://www.omim.org/	[76]
Chemical structure and Drug–target information	**Pharmacogenomics Knowledge Base (PharmGKB)**	The dataset of genetic variation on drug response	https://www.pharmgkb.org/	[77], [78]
Drug-disease information	**Pharos**	Resource connecting drugs, targets, and diseases	https://pharos.nih.gov/idg/index	[79]
Chemical structure	**PubChem**	Database of over 60 million compound structures, chemical features, bioactivity, etc.	http://pubchem.ncbi.nlm.nih.gov/	[80], [81]
Adverse effects and clinical trial information	**Side Effect Resource (SIDER)**	Information from public documents and package inserts on marketed compounds and their recorded ADEs, including side effect frequency	http://sideeffects.embl.de/	[82], [83]
Drug–target information	**STITCH**	Database of known and predicted chemical-protein interactions, which integrates the evidence derived from experiments, other databases and pieces of literature.	http://stitch.embl.de/	[84], [85]
Drug–target information	**The Binding Database (BindingDB)**	Information on binding affinities and other quantities related to Drug-target interactions	www.bindingdb.org/bind	[86], [87]
Drug omics data	**The Connectivity Map (CMap)**	Collection of 7,000 expression profiles genome-wide transcriptional expression data representing 1,309 compounds. Discovering patterns of association between drug sensitivity and gene expression signatures	www.broad.mit.edu/cmap	[88], [89]
Chemical structure Drug–target information	**Therapeutic Target Database (TTD)**	Provides the information about known and explored therapeutic protein and nucleic acid targets, the targeted diseases, pathway information and corresponding drugs directed at each of these targets.	http://bidd.nus.edu.sg/group/cjttd/	[90], [91]

review for drug repurposing. In this section, a triple analytical framework is introduced, whose effectiveness is indicated in selecting suitable methods and improvement paths for each approach. This paper aims to provide researchers with an analytical platform which can analyze the most common and recent methods and highlight the suitability and effectiveness of drug repurposing models by comparing them based on some evaluation criteria. Hence, the

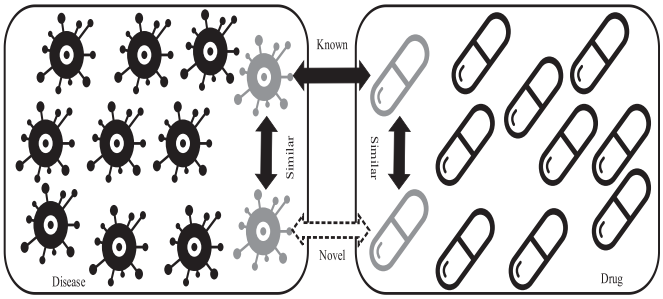

Fig. 3. Guilt-by-association: 'similar diseases' need the same therapies and can thus be treated with similar drugs [29].

structure of the proposed framework is established based on three segments:

1. Classification of computational drug repurposing models
2. Evaluation Criteria
3. Analytical assessment of frequent computational drug repurposing models

In the following, these three components are further elaborated.

4.1 Classification of Computational Drug Repurposing Models

In this section, an analytical review is presented, and its effectiveness in selecting a suitable method is illustrated. As can be seen in Fig. 4, we try to propose a classification for methods that can identify new relations between a set of drugs and diseases based on the known association between these two sets and similarities in each set (see Fig. 3).

An in-depth investigation of a large body of available research suggests that existing methods in computational drug repurposing can be categorized as follows Fig. 4. At the top level, our proposed classification can be divided into network-based approaches and learning-based approaches. Initially, in the network-based approach, datasets are integrated with network-driven data integration approach and benefit from advantages of graphs and networks [78], [99], [100]. Secondly, there are learning-based models which extract patterns from the input data integrated based on non-network driven data integration approaches [8], [28], [101], [102], [103], [104]. In each of the following methods, along with an elaborate description of each method, their main ideas, benefits, and drawbacks are also expressed.

4.1.1 Network-Based Approaches

This approach covers methods striving to find a new drug-disease relationship in an integrated network [3], [23]. Network-based methods offer some advantages: 1) Similarity; network topology measures, and clustering algorithms linked to topological properties of the network can be used to reinforce the hypothesis based on biological-based measures or to make predictions when biological information is missing [26]; 2) Analysis of topological features of the network such as communities and connectivity can itself lead to the discovery of new relations between compounds [26], [105]; 3) When the known drug-disease associations are organized into a drug-disease network, the drug repurposing can be replaced by link prediction problem in the complex network theory. Therefore, methods for drug repurposing can borrow ideas from the area of complex network analysis research [23]; 4) With the advances of high throughput technology and bioinformatics methods, molecular interactions in the biological systems can be modelled by networks [22]; 5) Network algorithms can readily accomplish tasks such as visualizing various existing interactions, adding newly discovered relationships, and superimposing additional properties over primary components and their known interactions [3]. These advantages make networked-based methods more popular than other methods these days. We categorized network-based methods into three main classes: link prediction, community detection, and graph-based semi-supervised.

- *Link prediction* is one of the essential tasks of link mining which is a part of data mining science [106]. Link prediction methods are those with the ability to recognize whether a connection exists between two objects or not [107]. Alternatively, they try to predict missing links in a network [108]. Link prediction can be defined as:

$$G = \langle V, E \rangle \; where \\ E_k = E_{k-1} \cup e_{new}, \tag{4}$$

Where, e_{new} is the new relation predicted in the network G (Fig. 6b).

For example, Wang et al. [51] proposed a novel heterogeneous network model which integrates drug repurposing and drug-target prediction problem into one unified framework. This network integrates 'diseases', 'drugs' and 'drug targets' (Fig. 5) as well as the drug repurposing formulated as a missing edge prediction problem on this heterogeneous

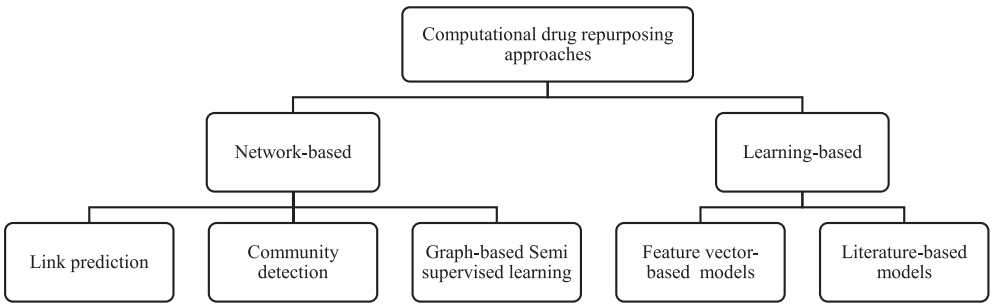

Fig. 4. The proposed classification of computational drug repurposing approaches.

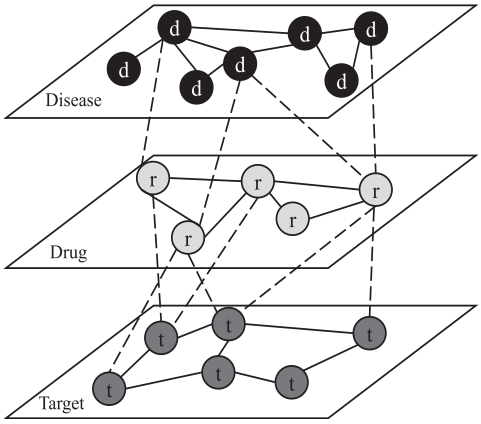

Fig. 5. A heterogeneous network model adopted from [51].

graph. Chen et al. [109] integrated and annotated data from public datasets relating to drugs, chemical compounds, protein targets, diseases, side effects and pathways, thereby developing a semantic linked network consisting of over 290,000 nodes and 720,000 edges. Then they developed a statistical model called Semantic Link Association Prediction (SLAP) to assess the association of drug target pairs and to predict missing links.

- *Community detection* in networks also called graph or network clustering [110], [111],aims to find a group of nodes and entities probably sharing common properties and/or playing similar roles within the network [112], with maximum connections within each group and relatively few between the groups [113]. The community detection can be defined as:

$$G = \bigcup_i G_i where \\ G_i = \{e_i \in G_i | \forall j \;\; e_{ij} > thershold\}, \tag{5}$$

Where, G involves the set of different subgraphs (community) where each relation in this subgraph has a stronger connection than a specific threshold (Fig. 6a).

The drug repurposing problem can be replaced by link prediction problem in the complex network theory. In other words, methods for drug repurposing can borrow ideas from the field of complex network analysis [23]. For example, Wu et al. [100] used known disease-gene and drug-target relationships, and then built a weighted disease and drug heterogeneous network. The nodes represented drugs or diseases while the edges refereed to shared gene, biological process, pathway, phenotype, or a combination of these features. Next, they clustered this weighted network to identify modules and then assembled all possible drug-disease pairs (putative drug repositioning candidates) from these modules. Udrescu et al. [115] presented a new approach based on clustering and topological community detection techniques for investigating drug-drug interaction networks. Their methodology revealed functional drug categories along with their intricate interrelationships. They linked the

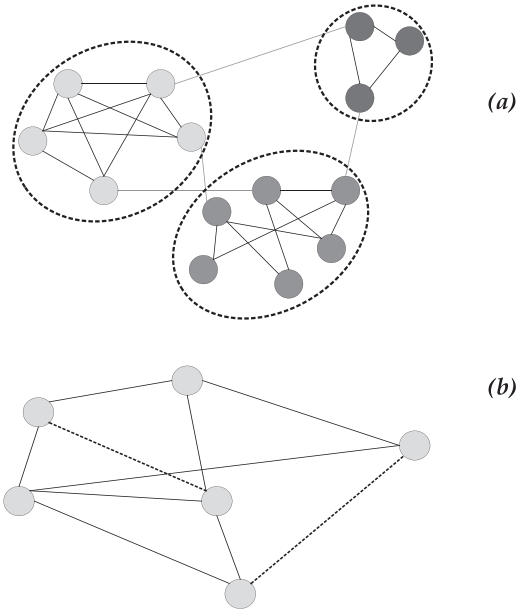

Fig. 6. A classic illustration of link prediction and community detection: (a) Community detection [114]. (b) Link prediction.

network clusters to nine relevant pharmacological properties by applying modularity-based and energy-model layout community detection algorithms. Using network centralities, they then ranked drugs according to their interaction potential for both simple and complex multi-pathology therapies.

- *Graph-based semi-supervised learning* algorithms, once the graph is created, learning will involve assigning label values to the vertices in the graph. This assignment becomes possible by edges that connect labelled vertices to unlabeled vertices. The graph edges are usually undirected.

$$x_i \neq x_j \\ if\, w_{ij} > \theta \, then \, y_i = y_j, \tag{6}$$

The edge between two vertices x_i and x_j represents the similarity of the two instances. Let w_{ij} be the edge weight. The idea is that if w_{ij} is large, then the two labels y_i and y_j are expected to be the same [116]. Nevertheless, in drug repurposing problems, labels will mostly be associated with edges instead of nodes (See Fig. 7).

Many different graph-based Semi-supervised learning algorithms, called label propagation [117] and graph cut [118], have been applied in drug repurposing. For example, Lotfi Shahreza et al. [52] proposed a novel semi-supervised heterogeneous label propagation algorithm (called *Heter-LP*) to find new interactions from a heterogeneous network consisting of information about drugs, diseases, and targets as collected from multiple sources at different levels. Le et al. [40] proposed a novel method called Regularized Least Square for Drug Repositioning(RLSDR) relying on a semi-supervised learning model, regularized least square. In this way, it did not require the definition of non-drug-disease associations as in

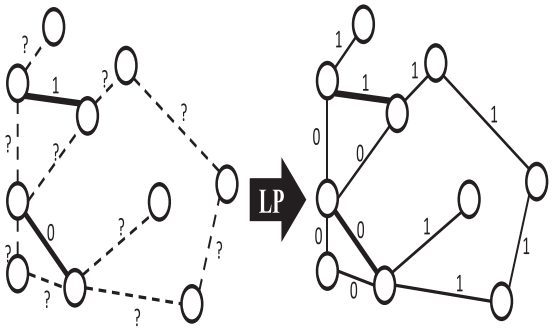

Fig. 7. Graph-based semi-supervised learning: label propagation.

previously proposed machine learning-based methods. Wu et al. [119] integrated heterogeneous data into three layers and constructed a novel weighted drug-disease pair network, where nodes were a drug-disease pair with known or unknown treatment relations and edges represented the node-node relations weighted with the similarity score between two pairs. Finally, they developed a semi-supervised graph cut algorithm (SSGC), to find the optimal graph cut, based on which they could identify the potential drug-disease treatment interactions.

4.1.2 Learning Based Approach

Learning-based approaches involve learning models presenting a model learned with a given feature vector and extract patterns, which eventually lead to the discovery of novel relations between drugs and diseases. Alternatively, they involve literature-based models in case we should deal with the biomedical and pharmaceutical knowledge available in literature or databases containing a vast volume of information [22], [120]. The main idea behind this grouping is based on identifying patterns in databases or natural language text to predict novel associations between drugs and targets or drugs and diseases [41]. The following two points describe this approach in more detail.

- *Feature vector-based models* are those methods that are based on machine learning methods which can be applied on different datasets related to drugs or diseases to predict therapeutic drug classes by available dataset's features constructed from similarity measures [26]. These methods are those that need the drug-disease pairs to be explicitly represented as fixed-length feature vectors [121].

  For example, Yamanishi et al. [122] investigated new interactions for four different drug-target classes using the Kernel Regression Method (KRM). Menden et al. [101] constructed machine-learning models to predict the response of specific cancer cell lines to drug therapy, quantified through IC50 values. This model used genomic cancer features of the cell lines and chemical properties to develop a feed-forward perceptron neural network model and a random forest regression model. Napolitano et al. [102] focused on a drug-centered approach to predict therapeutic drug classes using drug-related features. They joined these features into a single drug similarity matrix, which was adopted as a kernel for SVM classification. Later, Gottlieb et al. [8] in addition to drug-related features also integrated various disease-related features. Drug-drug and disease–disease similarity measures were computed to construct classification features based on these features, and then a logistic regression classifier was utilized to predict novel drug indications. Aliper et al. [123] indicated how deep neural networks (DNN) trained on sizeable transcriptional response data sets can classify multiple drugs to therapeutic categories totally based on their transcriptional profiles.

- *Literature-based models* can deal with biomedical and pharmaceutical knowledge available in the literature or databases containing a vast amount of information [22], [124]. Fig. 8 presents the general biomedical relation extraction workflow architecture that can be used in literature-based models. These methods rely on processing publicly accessible biomedical literature data to uncover indirect or innate relationships among seemingly unconnected biological entities [125], which can automatically mine [73], [74], [75], [129] and retrieve [130] novel drug-disease relations. Rastegar-Mojarad et al. [131] utilized drug-gene and gene-disease semantic predications extracted from Medline abstracts to create a list of potential drug-disease pairs. They then arranged the generated pairs by indicating scores based on the predicates qualifying drug-gene and gene-disease relationships. Having compared the top-ranked drug-disease pairs against

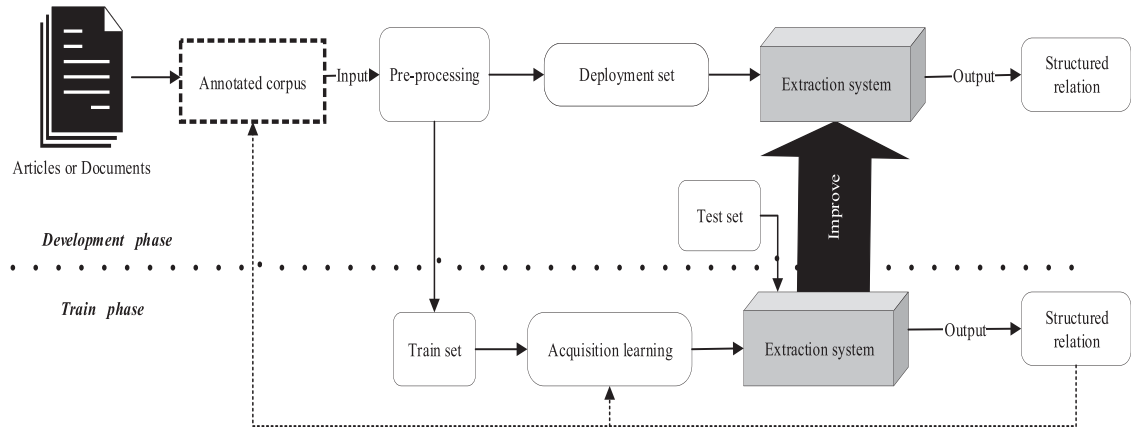

Fig. 8. Biomedical relation extraction workflow from [132].

TABLE 2
Benefits and Challenges of Computational Drug Repurposing Approaches

Approaches	Methods	Main idea	Benefits	Challenges
Network-based	**Link prediction**	Inferring the existence of unknown interactions between pairs of entities based on their properties and the currently observed links in the network [133].	1. Reducing the costs of empirical approaches	1. In heterogeneous networks dealing with multiple links and nodes may be intricate.
			2. Enable the extraction of absolute information existing in the network	2. The sparsity of linked data.
			3. Enable to modelling and evaluating network evolution mechanisms.	3. Choosing the right algorithm to perform the prediction on the available amount of information.
			4. Applicable and convenience in a variety of contexts	4. Size of the network can limit the kinds of techniques that can be applied.
	Community detection	Cluster networks to find a group of nodes and entity [112].	1. A community of vertices having a higher probability of being connected and have the same property than to members of other groups.	1. The quality of communities depends on their size.
			2. Can be used to create data structures to efficiently store the graph data and to handle navigational queries, like path searches.	2. The number of clusters for each method is different.
			3. Identifying modules and their boundaries allows for a classification of vertices, according to their basic position in the modules.	3. Since the clustering problem is ill-defined, there is no clear-cut answer that which algorithm shall we use on our data.
	Graph-based Semi-supervised learning	If the similarity of two nodes is large(nearby nodes), then their labels are expected to be the same [134].	1. Has a clear mathematical framework	1. High dependency to graph manufacturing(graph structure and weights)
			2. Has high performance in case of appropriate graph construction 3. Extendable to directional graphs 4. No need to identify negative association	2. Convergence of objective function
Learning based	**Feature vector-based models**	Methods that are based on machine learning methods on feature vectors [108]	1. Easy to implement and easy to understand	1. Necessity of constructing feature vector
			2. Just by having feature vector it can predict with any binary classification model	2. Lack of attention to the multi-relational structure of graph
	Literature-based models	These methods rely on the processing of publicly accessible biomedical literature data to uncover indirect or innate relationships among seemingly unconnected biological entities [125]	1. A grand deal of biological information is currently accessible in online text repositories such as Medline.	1. Need for adapting the natural language component to make it able to recognise the relevant entities and events correctly.
			2. Enable to discover associations, patterns and clusters of related texts.	2. Require a transposition of biomedical literature text into a structured form. 3. Require up-to-date information related to our favourite research topic.

the Comparative Toxicogenomics Database (CTD), they observed that a notable percentage of top-ranked pairs appeared in Comparative Toxicogenomics Database. Co-occurrence of these high-ranked pairs in Medline abstracts further increases the confidence in the approach to rank the inferred drug-disease relations higher in the list. Yang et al. [120] devised a pattern-based relationship extraction method to extract disease-gene and gene-drug direct relationships from the literature. These direct relationships are used to infer indirect relationships using the ABC model.

4.2 Evaluation Criteria

The second section of this framework is presented to introduce evaluation criteria in the computational drug

repurposing. Although in this paper *in silico* evaluation criteria are introduced, it can also be noted that in addition to these methods of evaluation, there are experimental methods which include *in vitro* and *in vivo* models (e.g., cell-based targeted assays and mouse models) These have been increasingly used to validate the candidate hits for preclinical drug evaluation [22].

Evaluation criteria represent a vital function in assessing the quality and validity of an algorithm. Concerning evaluation, it is important to mention that the evaluation criteria can be divided into quantitative and qualitative measures. Accordingly, in the following, the main quantitative and qualitative measures that can be useful in assessing the computational drug repurposing methods are introduced. This section is significantly fundamental due to its ability to examine the performance of a method proposed in computational drug repurposing.

- *Authenticity*. This criterion refers to a collection of all the quantitative metrics which are very important and variable in each study due to some serious challenges in computational evaluation of drug repurposing methods. One of the critical problems in quantitative assessment is benchmark data sets to which the ?algorithm is applied. In this regard, Brown and Patel [135], analyzed the types of databases that have been used for analytic validation in drug repurposing articles. According to them, although many of the investigators in the studies that they examined claimed to use a 'gold standard', there was substantial heterogeneity in the source of these standards as well as the types of data they contained. Further, for some criteria (like specificity) which require information about false positive (predicted drug indications that? are false), the researchers chose to mention all unannotated drug-indication pairs as false positives [135].

  As mentioned earlier, although there are some challenges, this is the only way to evaluate our results. Then, the quantitative evaluation of the results is measured using a set of metrics designed to assess the reliability and accuracy of the predictions. Finally, when a new method is implemented, or the already existing ones are modified, comparing the performances of the new method with those of already established ones using identical benchmark data sets is very helpful to understand to what extent and in which context the new method provides better predictions [26]. In this section, some qualitative criteria in computational drug repurposing are presented as a measure to study the authenticity of each approach. Although all of these qualitative criteria are not measurable due to the mentioned challenges, understanding their importance can help in the selection of a proper method for drug repurposing. The intention is to identify reasonable and practical qualification measures to evaluate computational drug repurposing. These metrics are typically the same for computational approaches, and can be divided into two broad categories: fixed-threshold metrics and threshold curves. All fixed-threshold metrics suffer the limitation that some estimates of a reasonable threshold must be available. On the other hand, threshold curve criteria such as ROC or PR are an alternative to these weaknesses [108], [136], [137].

  (a) *Fixed-threshold metrics* rely on diverse types of thresholds: prediction score, the percentage of instances, and the number of instances [137].

  - *Accuracy* is a commonly applied metrics in classification which can be defined as below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{7}$$

    Where, TP represents the true positive predicted relations while TN denotes the true negative predicted relations. Also, FP and FN refer to false positive and false negative relations, respectively. It represents the ability to predict the right class. To date, although some works are under development [135],such as repoDB [138] and RepurposeDB [139] (gold standard database), the field still lacks such a database which includes reliable true negatives because most of the times investigators chose to just mark all missing drug–indication pairs as true negatives [37], [135]. In addition, biological networks are sparse and existing links only constitute a small percentage of all possible links [136]; hence accuracy is not a meaningful measure.
  - *Specificity* measures the proportion of negative drug-disease relations that are correctly identified, which can be defined as below:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{8}$$

  - *Sensitivity* (recall) measures the proportion of positive drug-disease relations that are correctly identified, which can be defined as follows:

$$\text{Sensitivity} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

    Sensitivity assesses the general ability of repurposing methods to make reasonable claims [26], [135], rather than selecting a single or several high-ranking predictions to test in depth. Sensitivity is appealing because investigators only need to have a set of true positives to compare their predictions [135].
  - *Precision* is the fraction that shows the ratio of the true positive predictions among all positive predictions which can be defined as below:

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{10}$$

    It indicates the ability to discern biologically relevant interactions from untrue ones.

- *F1-score* is a harmonic mean of Precision and Recall, no matter what form of Precision and Recall is used. It can be employed to evaluate the overall performance, which is defined as:

$$F1\_score = 2 \times \frac{Recall \times Precision}{(Recall + Precison)}. \tag{11}$$

These three measures are commonly used as assessment indices in information extraction and machine learning [140]. In semantic models that we have to deal with text data, we can measure precision by:

$$\text{Precision} = \frac{|\{Relevant\} \cap \{Retrieved\}|}{|\{Retrieved\}|}. \tag{12}$$

And, recall by

$$\text{Recall} = \frac{|\{Relevant\} \cap \{Retrieved\}|}{|\{Relevant\}|}. \tag{13}$$

(b) *Threshold curves* are alternatives to fixed-threshold metrics due to the lack of incidents when researchers have a reasonable threshold at their disposal. Threshold curve works by shifting the threshold, computing metrics for each one, and then drawing a curve with all computed metrics in all thresholds. If the class distribution is very imbalanced, these curves become popular. Also, in the threshold curve, a single scalar measure known as area under the curve is used, which functions as a single summary statistic of performance cases when researchers are in possession of a reasonable threshold [108], [141]. Most methods evaluate their prediction performance using the area under the curve (AUC) and area under the precision-recall (AUPR) [3].

- *Receiver Operating Characteristics Curve*. (ROC) graphs are a useful technique for organizing classifiers and visualizing their performance. The ROC curve plots sensitivity vs (1 - specificity) of a test as the threshold fluctuates within its entire range. Each data point on the plot represents a distinct set of the threshold, and each threshold setting defines a particular set of TP, FP, TN, and FN counts, and consequently a specific pair of sensitivity and (1 - specificity) values. According to [142], the area under the curve (AUC), as a useful measure of accuracy, has been considered, with a meaningful AUC interpretation usually representing the overall performance of the algorithm [3]. AUC is comparable to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.
- *Precision-Recall Curve*. Previous studies have suggested that in scale-free networks, such as biological networks, PR curves are more informative because of the impact of skewed edge distributions on the performance of prediction algorithms [3], [23]. Precision-Recall curve is obtained from plotting precision against recall where the area under a PR curve is called AUPR.

- *Scalability*. This refers to the ability to construct the predictor efficiently given considerable amounts of data. As biological data mostly contain a bulk volume of information [143] and as the size of data increase the complexity dramatically, this has a diverse effect on scalability. Hence, existing methods should have the ability to deal with this.

$$\begin{aligned} Scalability &\propto Size^{-1} \\ Scalability &\propto Complexity^{-1}. \end{aligned} \tag{14}$$

- *The Ability to Face Heterogeneous and Complex Data*. With the explosive growth of heterogeneous data within bioinformatics, extracting valuable information from the vast amount of data is a challenging and vital task [55], [144]. Hence, the ability to face heterogeneous and complex data criteria refers to the ability of methods to handle biological data with a heterogeneous and complex nature in drug repurposing problem.
- *Straightforward*. Straightforward criteria capture to what extent these proposed methods are uncomplicated and easy to implement.
- *Robustness*. Robust methods in statistics and data mining are those resistant to the influence of noise and presence of outliers. This criterion (robustness) measures the ability of the predictor to make correct predictions given noisy data or data with missing values. It will also answer the question: "*Do the abilities of computational methods to deal with missing or noisy data differ* [145]?"
- *Complexity*. the computational complexity of an algorithm is analyzed as a function of its input size and to what extent it affects the length of time that methods consume. Feature, model, and parameter selection can have a significant impact on the performance of a data mining task (e.g., prediction). However, successfully optimizing each step requires overcoming considerable complexity which is difficult without in-depth knowledge and experience [146].
- *Generalization*. when a method has an appropriate result in response to most datasets, it has a great generalizability. This means that if a method uses less specific features and explores the features just by the input data, then its evaluation is highly independent of the selection of proper features [108], [143], [147], [148].

4.3 Analytical Assessment of Frequent Computational Drug Repurposing Models

The last and arguably the most important part of this analytical review is the evaluation of the computational drug

TABLE 3
Qualitative Evaluation Metrics for Assessment of Computational Drug Repurposing Models

Approach	Model	Authenticity	Scalability	Ability to face heterogeneous and complex data	Straight-forward	Robustness	Complexity	Generalization
Network-Based	Link prediction	High	Medium	High	Medium	High	High	Medium
	Community detection	High	Medium	High	Medium	Medium	High	Medium
	Graph-based Semi-Supervised Learning	High	Low	Low	Medium	Low	High	Medium
Learning-Based	Feature vector-based models	Medium	Low	Low	High	Medium	Medium	Low
	Literature-based models	Low	Medium	Medium	Low	Medium	High	Low

repurposing approach based on the metrics proposed in the 4.2 to assess to what extent these methods can satisfy these metrics, as summarized in Table 2. An attempt was made to show the efficiency of computational drug repurposing models in a different view to guide readers to understand their superiority over one another. It is noteworthy that this is a qualitative assessment based on a comprehensive study of the results of previous papers as well as the nature of methods in each approach; it is not a quantitative evaluation based on experimental trials. Every researcher has to face these questions:

- How can they compare each approach with each other?
- How should they select the best approach?
- What are the open avenues ahead of each approach for improvement?

In this paper, we present a discussion and comparison of the approaches with each other with a variety of evaluation criteria, and then we will explain how this analytical component can be used in answering the mentioned questions.

4.3.1 Comparison

As mentioned above, the last and possibly the most critical segment of the proposed framework is the evaluation of computational drug repurposing categorization methods. This segment evaluates computational drug repurposing techniques according to the criteria that have been introduced in the previous subsection. This qualitative comparison of the proposed approaches can be considered as a direction to a better understanding of the superiority of each method over the other. Our evaluation is summarized in Table 3. As observed in this table, the column headings represent the evaluation criteria, while the row headings show the nodes of the classification tree (Fig. 4).

The qualitative comparison of the proposed approaches is a guide to understanding their superiority over one another. Since the result of each method is based on different types of data sets and validation models, it should be noted that this evaluation is a qualitative assessment based on a precise study of the results of previous papers and the nature of each method, rather than a quantitative evaluation based on scientific experiments. The performance and prediction power of network-based methods vary depending on the similarities used and network sparsity. As the network size grows, the *authenticity* of network-based methods diminishes, as most real networks are large and sparse like biological networks [143], where with the growth of the network size, sparsity rises dramatically [149]. Generally, the authenticity of network-based methods improves with the amount of available data. However, since nearly almost every newly proposed network-based method has a higher authenticity than learning based methods, so medium authenticity is assigned to the learning models because authenticity is highly dependent on selecting the feature vector.

As discussed earlier, most real networks are large and sparse such as biological networks and when a network grows in size, it becomes mostly sparser than what it used to be [150]. In this case, it can be claimed that sparsity characteristic has a direct impact on *scalability*. Unfortunately, most semi supervised learning methods scale inadequately with the data size. They usually have a square time complexity $O(dn^2)$ (suppose that data lie in $\mathbb{R}^d$) for neighborhood graph construction and an approximately linear time complexity $O(kn)$ (k is a constant) for label propagation over a graph, so the final time complexity remains as $O(dn^2)$. Such a square time complexity is computationally prohibitive in large-scale applications, preventing the adoption of Graph based Semi Supervised Learning methods in practice [151].

Traditional feature vector-based models cannot *handle heterogeneous and complex datasets* effectively. In contrast, network-based methods are very useful to tackle this situation except for graph-based semi supervised learning. For example, most graph-based label propagation algorithms propagate label information only on a homogenous network, which is not suitable for spreading label information across heterogeneous networks [52].

The most *straightforward* tasks are based on feature vector-based models. Also, the validation procedure of this kind of method is more accessible according to the primary procedure itself. Concerning literature-based models, since they need a hard preprocessing procedure (e.g., NLP), we assume that literature-based models have the minimum straightforwardness.

The leaning-based models have typically less *generalizability*, but there is always some exception as in feature vector-based models, deep learning has a considerable generalizability. This is because they are very independent of the feature vectors and they extract latent features at another level of abstraction [108]. However, classical feature vector-based models have less generalizability, so we can say that generally learning-based models have low generalizability. Although many types of research have been conducted for scaling classifiers, the supervised model still has less generalizability. However, considering network-based methods, they are not very reliant on the selection of proper features; so as can be seen in Table 3 we consider their generalization as medium.

With regards to *robustness*, sparse networks are susceptible to noise [143]. So, due to this reason, network-based models have relatively low robustness.

In network-based models, as the size of the network grows, the computational *complexity* increases in response.

Further, the networks that are used in drug repurposing are almost huge, so the computational process might be very complex. As for learning-based models such as literature-based models, the main challenging issue arises from the complexity of a natural language itself, so literature-based models have a high complexity [152].

4.3.2 Selection

Another major purpose of our framework is preparation for choosing the effective approach at different levels of detail for use in computational drug repurposing. Since the drug repurposing approach is considered as an alternative to drug discovery, the main purpose would be finding the maximum number of potential interactions. It raises some basic questions as:

✓Which conditions result in discovering the maximum number of possible interactions between drugs and diseases?
✓Which drug repurposing approaches can be obtained to achieve the best result?

For answering these questions, it has been determined what technical models are acceptable for the computational drug repurposing by considering pros and cons of each approach (Table 2). Also Table 3 presents the assessment. As previously stated, computational drug repurposing can be divided into two categories: learning-based and network-based methods. This leads researchers to two distinct ways for selecting a suitable method for the repurposing problem. As shown in Table 3, for learning-based methods, literature-based models can be used for text format information and be considered as a biomedical relation extraction problem. The literature-based methods gather valuable, reliable, and affordable information from a growing number of free-form texts, though they suffer challenges including low authenticity and high complexity. On the other hand, supervised models are very straightforward with acceptable complexity, although their ability to face heterogeneous data and scalability remains a challenge to deal with. Network-based data are very useful to handle heterogeneous and complex data sets, but the complexity of these methods is too high. Link prediction methods face the problem of dependence of the type of technique which can be applied on the size of the network, while these methods have a great authenticity and resistance to noise. The power of community detection methods is also very dependent on the size of the network. Graph-based Semi-Supervised learning approach is highly sensitive to graph structure and weights, but their result is highly accurate.

4.3.3 Improvement

In addition to assessing and identifying the index criteria in selecting the basic approach, making significant improvements on methods is another major aim of this assessment. Our framework provides an overview of prospects of each method which can be achieved by combining and resolving the challenges of computational drug repurposing methods.

- *Combination* Table 3 can be used explicitly for combining methods. Notably, a combination of several methods is often more desired for achieving better results in that this combination has revealed better performance when compared with individual methods. As an instance, use of a different type of data integration approach and then applying analytical approaches yield better results [41], [96]. For example, [153] integrated data with kernel-based methods and then applied learning-based approach (i.e., supervised: SVM) to identify drug-disease relationships, and further made a drug-disease network-based analysis. The results obtained indicated that the two methods combined were more sensitive than when applied individually. So our framework could be a source for understanding how models can be combined to mitigate the shortcomings of each other. We can guess which models are compatible with each other based on the main ideas in Tables 2 and 3. Indeed, they can help to determine whether the combination of which models can improve their low or medium degrees. In other words, it can help to discuss if the combination of models from the same or different approach can be beneficial or not. In addition, Table 3 could be very valuable for integrating various models which can lead researchers to introduce a more desirable model and becoming a guide to further research efforts. On the other hand, standard repurposing algorithms that are supposed to make predictions for candidates for which no interaction is known have some limits; so to overcome this limitation, existing methods must be adapted. For instance, [154] used the findings of paper [8] who collected comprehensive associations between drugs and diseases from multiple data sources to infer secondary associations between drugs and diseases based on link prediction and community detection methods.
- *Challenge elimination* of our framework provides researchers with the ability to reduce the challenges of each approach. Since an important matter in drug development is the discovery of novel drug targets, other associations and interactions related to drug development can also be used in drug repurposing; so with the integration of these sources more reliable results can be obtained. Although challenges in drug repurposing still hinders us from practically using repurposed drugs, some measures can be helpful in finding better practical models for the drug repurposing problem. These measures include continuing research in this field and related subjects such as predicting adverse drug reactions and drug-target relations as well as analyzing the proposed models in these other subjects and benefitting from their advantages and disadvantages and results of their work,. One of the essential challenges in drug repurposing is variations in the types and sources of annotation data which can be used for performing validation results in light of instability in the field. Researchers can tackle this challenge by creating a real 'gold standard' which leads to both repurposing successes and failures and allows for fair comparisons between methods. Accordingly, it is reasonable to argue that such a 'gold standard' database can enhance the accuracy of drug repurposing methods and raise the probability of success in clinical trials. Notably, since the proposed methods up to now

have been different considering the model validation and the purpose of the evaluation criteria, the analytical assessment in Table 3 can be beneficial.

- *Future ideas* for computational drug repurposing can originate from some facts presented in this review. This review helps researchers understand that computational drug repurposing models lack some trending subjects including deep learning algorithms [155] and use of different types of databases such as electronic health records [156], [157]. Further, the represented classification of methods in this review leads to some open avenues for further investigations in each approach, for finding new methods and ideas. For example, a glance at the Table 3 reveals that link prediction methods are powerful and have the potential to be ameliorated in future. This potential can be explored in link prediction methods represented by other complex networks analysis researchers. In this way, researchers in computational drug repurposing can find out about the benefits of using matrix factorization methods as a more accurate way of link prediction rather than random walk methods [158], [159], [160].

We believe that by closely studying Tables 2 and 3, many new paths and possibly improvements can be obtained in the field of computational drug repurposing.

5 Conclusion

With the full exploitation of data mining and biological science, computational drug repurposing strives to find new uses for existing drugs thereby making a huge difference in the process of health promotion. Here, we reviewed the current, state-of-the-art, and most common methods that have been used in many drug repurposing studies. Then, we presented a systematic analysis of computational drug repurposing. We described the advantages and disadvantages of every approach of modelling drug repurposing problem to develop better models, despite the inevitable challenges and considering available data sources. In this paper, a framework was introduced involving a new classification for computational drug repurposing methods from a different view and with an artificial intelligence perspective. In addition, a collection of the most current and notable works was included. Finally, a novel assessment of this categorization based on the presented evaluation criteria was presented which can be used to direct future studies in drug repurposing for comparison, selection, and improvement of each computational drug repurposing approach.

References

[1] S. J. Barrett and W. B. Langdon, "Advances in the application of machine learning techniques in drug discovery, design and development," in *Applications of Soft Computing. Advances in Intelligent and Soft Computing*, A. Tiwari, R. Roy, J. Knowles, E. Avineri, and K. Dahal, eds., Berlin, Heidelberg: Springer, 2006, doi: 10.1007/978-3-540-36266-1_10.

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

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

[4] S. Pushpakom, et al., "Drug repurposing: Progress, challenges and recommendations," *Nature Rev. Drug Discovery*, vol. 18, no. 1, pp. 41–58, Jan. 2019, doi: 10.1038/nrd.2018.168.

[5] M. M. Mehndiratta, S. Wadhai, B. Tyagi, N. Gulati, and M. Sinha, "Drug repositioning," *Int. J. Epilepsy*, vol. 3, no. 2, pp. 91–94, Dec. 2016, doi: 10.1016/j.ijep.2016.09.002.

[6] Y. Y. Wang, C. Cui, L. Qi, H. Yan, and X. M. Zhao, "DrPOCS: Drug repositioning based on projection onto convex sets," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 1, pp. 154–162, Jan./Feb. 2019, doi: 10.1109/TCBB.2018.2830384.

[7] K. Shameer, B. Readhead, and J. T. Dudley, "Computational and experimental advances in drug repositioning for accelerated therapeutic stratification," *Current Topic Med. Chem.*, vol. 15, no. 1, pp. 5–20, 2015, doi: 10.2174/1568026615666150112103510.

[8] A. Gottlieb, G. Y. Stein, E. Ruppin, and R. Sharan, "PREDICT: A method for inferring novel drug indications with application to personalized medicine," *Mol. Syst. Biol.*, vol. 7, no. 1, pp. 496–496, 2014, doi: 10.1038/msb.2011.26.

[9] S. Ekins, A. J. Williams, M. D. Krasowski, and J. S. Freundlich, "In silico repositioning of approved drugs for rare and neglected diseases," *Drug Discovery Today*, vol. 16, no. 7–8, pp. 298–310, Apr. 2011, doi: 10.1016/j.drudis.2011.02.016.

[10] R. Muthyala, "Orphan/rare drug discovery through drug repositioning," *Drug Discovery Today Therapeutic Strategies*, vol. 8, no. 3–4, pp. 71–76, Dec. 2011, doi: 10.1016/j.ddstr.2011.10.003.

[11] B. Delavan, R. Roberts, R. Huang, W. Bao, W. Tong, and Z. Liu, "Computational drug repositioning for rare diseases in the era of precision medicine," *Drug Discovery Today*, vol. 23, no. 2, pp. 382–394, 2018, doi: 10.1016/j.drudis.2017.10.009.

[12] D. Sardana, C. Zhu, M. Zhang, R. C. Gudivada, L. Yang, and A. G. Jegga, "Drug repositioning for orphan diseases," *Briefings Bioinf.*, vol. 12, no. 4, pp. 346–356, 2011, doi: 10.1093/bib/bbr021.

[13] E. Almeida, M. De Oliveira, and K. L. Lang, "Drug repositioning: Concept, classification, methodology, and importance in rare/orphans and neglected diseases," *J. Appl. Pharmaceutical Sci.*, vol. 8, no. 8, pp. 157–165, 2018, doi: 10.7324/JAPS.2018.8822.

[14] G. Jin, C. Fu, H. Zhao, K. Cui, J. Chang, and S. T. C. Wong, "A novel method of transcriptional response analysis to facilitate drug repositioning for cancer therapy," *Cancer Res.*, vol. 72, no. 1, pp. 33–44, 2012, doi: 10.1158/0008-5472.CAN-11-2333.

[15] G. J. Yoshida, "Therapeutic strategies of drug repositioning targeting autophagy to induce cancer cell death: From pathophysiology to treatment," *J. Hematology Oncology*, vol. 10, 2017, Art. no. 67, doi: 10.1186/s13045-017-0436-9.

[16] Y. Lo, S. Senese, B. France, A. A. Gholkar, R. Damoiseaux, and J. Z. Torres, "Computational cell cycle profiling of cancer cells for prioritizing FDA-approved drugs with repurposing potential," *Sci. Rep.*, vol. 7, 2017, Art. no. 11261, doi: 10.1038/s41598-017-11508-2.

[17] W. Younis, S. Thangamani, and M. N. Seleem, "Repurposing non-antimicrobial drugs and clinical molecules to treat bacterial infections," *Curr. Pharmaceutical Des.*, vol. 21, no. 28, pp. 4106–4111, 2015, doi: 10.2174/1381612821666150506154434.

[18] P. Balasundaram, S. Veerappapillai, S. Krishnamurthy, and R. Karuppasamy, "Drug repurposing: An approach to tackle drug resistance in S. typhimurium," *J. Cell. Biochem.*, vol. 119, no. 3, pp. 2818–2831, Mar. 2018, doi: 10.1002/jcb.26457.

[19] A. Maitra, S. Bates, T. Kolvekar, P. V. Devarajan, J. D. Guzman, and S. Bhakta, "Repurposing—a ray of hope in tackling extensively drug resistance in tuberculosis," *Int. J. Infectious Diseases*, vol. 32, pp. 50–55, Mar. 2015, doi: 10.1016/j.ijid.2014.12.031.

[20] S. Alaimo, R. Giugno, and A. Pulvirenti, "Recommendation techniques for drug–target interaction prediction and drug repositioning," in *Data Mining Techniques for the Life Sciences*, Louisville, Kentucky, USA: Humana Press, vol. 2017, no. 16, 2016, pp. 441–462, doi: 10.1007/978-1-4939-3572-7_23.

[21] L. Huang, et al., "DrugComboRanker: Drug combination discovery based on target network analysis," *Bioinf.*, vol. 30, no. 12, pp. 228–236, 2014, doi: 10.1093/bioinformatics/btu278.

[22] J. Li, S. Zheng, B. Chen, A. J. Butte, S. J. Swamidass, and Z. Lu, "A survey of current trends in computational drug repositioning," *Briefings Bioinf.*, vol. 17, no. 1, pp. 2–12, Jan. 2016, doi: 10.1093/bib/bbv020.

[23] Z. Wu, Y. Wang, and L. Chen, "Network-based drug repositioning," *Mol. Biosystems*, vol. 9, no. 6, 2013, Art. no. 1268, doi: 10.1039/c3mb25382a.
[24] J. S. Shim and J. O. Liu, "Recent advances in drug repositioning for the discovery of new anticancer drugs," *Int. J. Biol. Sci.*, vol. 10, no. 7, pp. 654–663, 2014, doi: 10.7150/ijbs.9224.
[25] R. A. Hodos, B. A. Kidd, K. Shameer, B. P. Readhead, and J. T. Dudley, "In silico methods for drug repurposing and pharmacology," *Wiley Interdisciplinary Rev. Syst. Biol. Med.*, vol. 8, no. 3, pp. 186–210, May 2016, doi: 10.1002/wsbm.1337.
[26] Q. Vanhaelen, et al., "Design of efficient computational workflows for in silico drug repurposing," *Drug Discovery Today*, vol. 22, no. 2, pp. 210–222, Feb. 2017, doi: 10.1016/j.drudis.2016.09.019.
[27] E. March-Vila, et al.., "On the integration of in silico drug design methods for drug repurposing," *Front. Pharmacology*, vol. 8, pp. 1–7, May 2017, Art. no. 298, doi: 10.3389/fphar.2017.00298.
[28] P. Zhang, F. Wang, and J. Hu, "Towards drug repositioning: A unified computational framework for integrating multiple aspects of drug similarity and disease similarity," in *Proc. AMIA Annu. Symp.*, vol. 2014, 2014, pp. 1258–1267.
[29] S. S. Sadeghi and M. Keyvanpour, "Computational drug repurposing: Classification of the research opportunities and challenges," *Current Comput. Aided Drug Des.*, early access, 2019, doi: 10.2174/1573409915666190613113822.
[30] A. Talevi, "Drug repositioning: Current approaches and their implications in the precision medicine era," *Expert Rev. Precision Med. Drug Develop.*, vol. 3, no. 1, pp. 49–61, 2018, doi: 10.1080/23808993.2018.1424535.
[31] T.-W. Kim, "Drug repositioning approaches for the discovery of new therapeutics for Alzheimer's disease," *Neurotherapeutics*, vol. 12, no. 1, pp. 132–142, Jan. 2015, doi: 10.1007/s13311-014-0325-7.
[32] N. Novac, "Challenges and opportunities of drug repositioning," *Trends Pharmacological Sci.*, vol. 34, no. 5, pp. 267–272, May 2013, doi: 10.1016/j.tips.2013.03.004.
[33] W. Sun, P. E. Sanderson, and W. Zheng, "Drug combination therapy increases successful drug repositioning," *Drug Discovery Today*, vol. 21, no. 7, pp. 1189–1195, 2016, doi: 10.1016/j.drudis.2016.05.015.
[34] J. T. Dudley, T. Deshpande, and A. J. Butte, "Exploiting drug-disease relationships for computational drug repositioning," *Briefings Bioinf.*, vol. 12, no. 4, pp. 303–311, Jul. 2011, doi: 10.1093/bib/bbr013.
[35] J. Zou, M. Zheng, G. Li, and Z. Su, "Advanced systems biology methods in drug discovery and translational biomedicine.," *Biomed. Res. Int.*, vol. 2013, 2013, Art. no. 8, doi: http://dx.doi.org/10.1155/2013/742835.
[36] E. Sam and P. Athri, "Web-based drug repurposing tools: A survey," *Briefings Bioinf.*, vol. 20, no. 1, pp. 299–316, Jan. 2019, doi: 10.1093/bib/bbx125.
[37] J. Yella, S. Yaddanapudi, Y. Wang, and A. Jegga, "Changing trends in computational drug repositioning," *Pharmaceuticals*, vol. 11, no. 2, Jun. 2018, Art. no. 57, doi: 10.3390/ph11020057.
[38] H. Xue, J. Li, H. Xie, and Y. Wang, "Review of drug repositioning approaches and resources," *Int. J. Biol. Sci.*, vol. 14, no. 10, pp. 1232–1244, 2018, doi: 10.7150/ijbs.24612.
[39] J. Langedijk, A. K. Mantel-Teeuwisse, D. S. Slijkerman, and M. H. D. B. Schutjens, "Drug repositioning and repurposing: Terminology and definitions in literature," *Drug Discovery Today*, vol. 20, no. 8, pp. 1027–1034, Aug. 2015, doi: 10.1016/j.drudis.2015.05.001.
[40] D.-H. Le and D. Nguyen-Ngoc, "Drug repositioning by integrating known disease-gene and drug-target associations in a semi-supervised learning model," *Acta Biotheoretica*, vol. 66, no. 4, pp. 315–331, Dec. 2018, doi: 10.1007/s10441-018-9325-z.
[41] M. Alshahrani and R. Hoehndorf, "Drug repurposing through joint learning on knowledge graphs and literature," *bioRxiv*, 2018, doi: 10.1101/385617.
[42] H. Moghadam, M. Rahgozar, and S. Gharaghani, "Scoring multiple features to predict drug disease associations using information fusion and aggregation," *SAR QSAR Environ. Res.*, vol. 27, no. 8, pp. 609–628, Aug. 2016, doi: 10.1080/1062936X.2016.1209241.
[43] M. Lotfi Shahreza, N. Ghadiri, S. R. Mousavi, J. Varshosaz, and J. R. Green, "A review of network-based approaches to drug repositioning," *Briefings Bioinf.*, vol. 19, no. 5, pp. 878–892, Sep. 2018, doi: 10.1093/bib/bbx017.
[44] H. Li, et al., "TarFisDock: A web server for identifying drug targets with docking approach," *Nucl. Acids Res.*, vol. 34, no. WEB. SERV. ISS., pp. 219–224, 2006, doi: 10.1093/nar/gkl114.
[45] X. Zhou, et al., " In silico drug repositioning based on drug-miRNA associations," *Briefings Bioinf.*, 2019, Art. no. bbz012, doi: 10.1093/bib/bbz012.
[46] Y. Murakami, L. P. Tripathi, P. Prathipati, and K. Mizuguchi, "Network analysis and in silico prediction of protein–protein interactions with applications in drug discovery," *Current Opinion Struct. Biol.*, vol. 44, pp. 134–142, 2017, doi: 10.1016/j.sbi.2017.02.005.
[47] J. Mullen, S. J. Cockell, P. Woollard, and A. Wipat, "An integrated data driven approach to drug repositioning using gene-disease associations," *PLoS One*, vol. 11, no. 5, pp. 1–24, 2016, doi: 10.1371/journal.pone.0155811.
[48] E. W. Su and T. M. Sanger, "Systematic drug repositioning through mining adverse event data in ClinicalTrials.gov," *PeerJ*, vol. 5, 2017, Art. no. e3154, doi: 10.7717/peerj.3154.
[49] O. Serçinoğlu and P. O. Sarica, "In Silico databases and tools for drug repurposing," in *In Silico Drug Design: Repurposing Techniques and Methodologies*, 1st ed., K. Roy Ed., Amsterdam, Netherlands: Elsevier Inc., 2019, pp. 703–742, doi: 10.1016/B978-0-12-816125-8.00024-9.
[50] S. Sarvagalla, S. B. Syed, and M. S. Coumar, "An overview of computational methods, tools, servers, and databases for drug repurposing," in *In Silico Drug Design*, Amsterdam, Netherlands: Elsevier Inc., 2019, pp. 743–780, doi: 10.1016/B978-0-12-816125-8.00025-0.
[51] W. Wang, S. Yang, X. Zhang, and J. Li, "Drug repositioning by integrating target information through a heterogeneous network model," *Bioinf.*, vol. 30, no. 20, pp. 2923–2930, 2014, doi: 10.1093/bioinformatics/btu403.
[52] M. Lotfi Shahreza, N. Ghadiri, S. R. Mousavi, J. Varshosaz, and J. R. Green, "Heter-LP: A heterogeneous label propagation algorithm and its application in drug repositioning," *J. Biomed. Inf.*, vol. 68, pp. 167–183, Apr. 2017, doi: 10.1016/j.jbi.2017.03.006.
[53] H. Luo, et al., "Computational drug repositioning with random walk on a heterogeneous network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, to be published, doi: 10.1109/TCBB.2018.2832078.
[54] V. Martínez, C. Navarro, C. Cano, W. Fajardo, and A. Blanco, "DrugNet: Network-based drug–disease prioritization by integrating heterogeneous data," *Artif. Intell. Med.*, vol. 63, no. 1, pp. 41–49, 2015, doi: 10.1016/j.artmed.2014.11.003.
[55] C. Goble and R. Stevens, "State of the nation in data integration for bioinformatics," *J. Biomed. Inform.*, vol. 41, no. 5, pp. 687–693, 2008, doi: 10.1016/j.jbi.2008.01.008.
[56] Y. Li, F.-X. Wu, and A. Ngom, "A review on machine learning principles for multi-view biological data integration," *Briefings Bioinf.*, vol. 19, no. 2, pp. 325–340, Mar. 2018. [Online]. Available: https://doi.org/10.1093/bib/bbw113
[57] A. Brazma, et al., "ArrayExpress - A public repository for microarray gene expression data at the EBI," *Nucl. Acids Res.*, vol. 31, no. 1, pp. 68–71, 2003, doi: 10.1093/nar/gkg091.
[58] N. Kolesnikov, et al., "ArrayExpress update-simplifying data submissions," *Nucl. Acids Res.*, vol. 43, no. D1, pp. D1113–D1116, 2015, doi: 10.1093/nar/gku1057.
[59] J. Barretina, et al., "The cancer cell line encyclopedia enables predictive modeling of anticancer drug sensitivity," *Nature*, vol. 483, no. 7391, pp. 603–607, 2012, doi: 10.1038/nature11003.The.
[60] A. Gaulton, et al., "ChEMBL: A large-scale bioactivity database for drug discovery," *Nucl. Acids Res.*, vol. 40, no. D1, pp. 1100–1107, 2012, doi: 10.1093/nar/gkr777.
[61] A. Gaulton, et al., "The ChEMBL database in 2017," *Nucl. Acids Res.*, vol. 45, no. D1, pp. D945–D954, 2017, doi: 10.1093/nar/gkw1074.
[62] A. P. Davis, et al., "The comparative toxicogenomics database: Update 2019," *Nucl. Acids Res.*, vol. 47, no. D1, pp. D948–D954, 2019, doi: 10.1093/nar/gky868.
[63] Y. Liu, Q. Wei, G. Yu, W. Gai, Y. Li, and X. Chen, "DCDB 2.0: A major update of the drug combination database," *Database*, vol. 2014, pp. 1–6, 2014, doi: 10.1093/database/bau124.
[64] C. Pacini, et al., "DvD: An R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data," *Bioinf.*, vol. 29, no. 1, pp. 132–134, 2013, doi: 10.1093/bioinformatics/bts656.
[65] E. J. Lo, et al., "DrugBank 5.0: A major update to the DrugBank database for 2018," *Nucl. Acids Res.*, vol. 46, no. D1, pp. D1074–D1082, 2017, doi: 10.1093/nar/gkx1037.

[66] D. S. Wishart, "DrugBank: A comprehensive resource for in silico drug discovery and exploration," *Nucl. Acids Res.*, vol. 34, no. 90001, pp. D668–D672, 2006, doi: 10.1093/nar/gkj067.

[67] D. Szklarczyk, et al., "The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible," *Nucl. Acids Res.*, vol. 45, no. D1, pp. D362–D368, 2017, doi: 10.1093/nar/gkw937.

[68] D. Szklarczyk, et al., "STRING v11 : Protein – protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets," vol. 47, no. D1, pp. D607–D613, 2019, doi: 10.1093/nar/gky1131.

[69] K. Hosomi, M. Fujimoto, K. Ushio, L. Mao, J. Kato, and M. Takada, "An integrative approach using real-world data to identify alternative therapeutic uses of existing drugs," *PLoS One*, vol. 13, no. 10, pp. 1–19, 2018, doi: 10.1371/journal.pone.0204648.

[70] O. Ursu, et al., "DrugCentral 2018 : An update," vol. 47, no. October 2018, pp. 963–970, 2019, doi: 10.1093/nar/gky963.

[71] T. Barrett, et al., "NCBI GEO: Mining tens of millions of expression profiles - Database and tools update," *Nucl. Acids Res.*, vol. 35, no. SUPPL. 1, pp. 562–566, 2007, doi: 10.1093/nar/gkl887.

[72] T. Barrett, et al., "NCBI GEO : archive for functional genomics data sets — update," vol. 41, no. D1, pp. D991–D995, 2013, doi: 10.1093/nar/gks1193.

[73] M. Qi, R. Wang, B. Jing, F. Jian, C. Ning, and L. Zhang, "KEGG: Kyoto encyclopedia of genes and genomes," *Infection Genetics Evol.*, vol. 44, no. 1, pp. 313–317, 2016, doi: 10.1016/j.meegid.2016.07.022.

[74] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, "KEGG for integration and interpretation of large-scale molecular data sets," *Nucl. Acids Res.*, vol. 40, no. D1, pp. D109–D114, Jan. 2012, doi: 10.1093/nar/gkr988.

[75] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions," *Sci. Transl. Med.*, vol. 4, no. 125, 2012, Art. no. 125ra31, doi: 10.1126/scitranslmed.3003377.

[76] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, and V. A. McKusick, "Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders," *Nucl. Acids Res.*, vol. 33, no. SUPPL. 1, pp. D514–D517, 2005, doi: 10.1093/nar/gki033.

[77] C. F. Thorn, T. E. Klein, and R. B. Altman, "PharmGKB: The pharmacogenomics knowledge base," *Methods Mol Biol.* vol. 1015, pp. 311–320, 2013, doi: 10.1007/978-1-62703-435-7.

[78] Z. Wu, W. Li, G. Liu, and Y. Tang, "Network-based methods for prediction of drug-target interactions," *Front. Pharmacology*, vol. 9, 2018, Art no. 1134, doi: 10.3389/fphar.2018.01134.

[79] A. Simeonov, et al., "Pharos: Collating protein information to shed light on the druggable genome," *Nucl. Acids Res.*, vol. 45, no. D1, pp. D995–D1002, 2016, doi: 10.1093/nar/gkw1072.

[80] J. Chen, et al., "PubChem substance and compound databases," *Nucl. Acids Res.*, vol. 44, no. D1, pp. D1202–D1213, 2015, doi: 10.1093/nar/gkv951.

[81] Y. Wang, et al., "PubChem BioAssay: 2017 update," *Nucl. Acids Res.*, vol. 45, no. D1, pp. D955–D963, 2017, doi: 10.1093/nar/gkw1118.

[82] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, "The SIDER database of drugs and side effects," *Nucl. Acids Res.*, vol. 44, no. D1, pp. D1075–D1079, 2016, doi: 10.1093/nar/gkv1075.

[83] J. Von Eichborn, M. S. Murgueitio, M. Dunkel, S. Koerner, P. E. Bourne, and R. Preissner, "PROMISCUOUS: A database for network-based drug-repositioning," *Nucl. Acids Res.*, vol. 39, no. SUPPL. 1, pp. 1060–1066, 2011, doi: 10.1093/nar/gkq1037.

[84] M. Kuhn, C. von Mering, M. Campillos, L. J. Jensen, and P. Bork, "STITCH: Interaction networks of chemicals and proteins," *Nucl. Acids Res.*, vol. 36, no. SUPPL. 1, pp. 684–688, 2008, doi: 10.1093/nar/gkm795.

[85] 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," *Nucl. Acids Res.*, vol. 44, no. D1, pp. D380–D384, 2016, doi: 10.1093/nar/gkv1277.

[86] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson, "BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities," *Nucl. Acids Res.*, vol. 35, no. SUPPL. 1, pp. 198–201, 2007, doi: 10.1093/nar/gkl999.

[87] 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," *Nucl. Acids Res.*, vol. 44, no. D1, pp. D1045–D1053, 2016, doi: 10.1093/nar/gkv1072.

[88] A. Subramanian, et al., "A next generation connectivity map: L1000 platform and the first 1,000,000 profiles," *Cell*, vol. 171, no. 6, pp. 1437–1452, 2018, doi: 10.1016/j.cell.2017.10.049.A.

[89] A. Musa, et al., "A review of connectivity map and computational approaches in pharmacogenomics," *Briefings Bioinf.*, vol. 19, no. 3, pp. 506–523, 2018, doi: 10.1093/bib/bbw112.

[90] X. Chen, "TTD: Therapeutic target database," *Nucl. Acids Res.*, vol. 30, no. 1, pp. 412–415, 2002, doi: 10.1093/nar/30.1.412.

[91] Y. H. Li, et al., "Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics," *Nucl. Acids Res.*, vol. 46, no. D1, pp. D1121–D1127, 2018, doi: 10.1093/nar/gkx1076.

[92] R. Wlison, *Introduction to GraphTheory* (4th Edition). Harlow, United Kingdom: Longman group Ltd, 1996.

[93] V. Gligorijević and N. Pržulj, "Methods for biological data integration: perspectives and challenges," *J. Roy Soc. Interface*, vol. 12, no. 112, pp. 1–19, Nov. 2015, doi: 10.1098/rsif.2015.0571.

[94] V. Gligorijevic, "Computational methods for integration of biological data," *Personalized Med.*, vol. 2, pp. 137–178, 2016, doi: 10.1007/978-3-319-39349-0.

[95] H. Lee, S. Kang, and W. Kim, "Drug repositioning for cancer therapy based on large-scale drug-induced transcriptional signatures," *PLoS One*, vol. 11, no. 3, pp. 1–17, 2016, doi: 10.1371/journal.pone.0150460.

[96] C.-H. Huang, P. M.-H. Chang, C.-W. Hsu, C.-Y. F. Huang, and K.-L. Ng, "Drug repositioning for non-small cell lung cancer by using machine learning algorithms and topological graph theory," *BMC Bioinf.*, vol. 17, no. S1, 2016, Art. no. S2, doi: 10.1186/s12859-015-0845-0.

[97] E. T. Shah, et al., "Repositioning 'old' drugs for new causes: identifying new inhibitors of prostate cancer cell migration and invasion," *Clin. Exp. Metastasis*, vol. 33, no. 4, pp. 385–399, 2016, doi: 10.1007/s10585-016-9785-y.

[98] C. Ng, R. Hauptman, Y. Zhang, P. E. Bourne, and L. Xie, "Anti-infectious drug repurposing using an integrated chemical genomics and structural systems biology approach," in *Proc. Pacific Symp. Biocomputing*, 2014, vol. 2014, pp. 136–147, doi: 10.1142/9789814583220_0014.

[99] J. Li and Z. Lu, "Pathway-based drug repositioning using causal inference," *BMC Bioinf.*, vol. 14, no. Suppl 16, 2013, Art. no. S3, doi: 10.1186/1471-2105-14-S16-S3.

[100] C. Wu, R. C. Gudivada, B. J. Aronow, and A. G. Jegga, "Computational drug repositioning through heterogeneous network clustering," *BMC Syst. Biol.*, vol. 7, no. Suppl 5, 2013, Art. no. S6, doi: 10.1186/1752-0509-7-S5-S6.

[101] M. P. Menden, et al., "Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties," *PLoS One*, vol. 8, no. 4, 2013, Art. no. e61318, doi: 10.1371/journal.pone.0061318.

[102] F. Napolitano, et al., "Drug repositioning: A machine-learning approach through data integration," *J. Cheminform.*, vol. 5, 2013, Art. no. 30, doi: 10.1186/1758-2946-5-30.

[103] Q. Zhu, C. Tao, F. Shen, and C. G. Chute, "Exploring the pharmacogenomics knowledge base (PharmGKB) for repositioning breast cancer drugs by leveraging Web ontology language (OWL) and cheminformatics approaches," in *Proc. Pacific Symp. Biocomputing*, 2014, pp. 172–182.

[104] C. Andronis, A. Sharma, V. Virvilis, S. Deftereos, and A. Persidis, "Literature mining, ontologies and information visualization for drug repurposing," *Briefings Bioinf.*, vol. 12, no. 4, pp. 357–368, 2011, doi: 10.1093/bib/bbr005.

[105] S. S. Sadeghi and M. Keyvanpour, "RCDR: A recommender based method for computational drug repurposing," in *Proc. 5th Conf. Knowl. Based Eng. Innovation*, 2019, pp. 467–471, doi: 10.1109/KBEI.2019.8734933.

[106] L. Getoor and C. P. Diehl, "Link mining: A survey," *ACM SIGKDD Explor. Newslett.*, vol. 7, no. 2, pp. 3–12, 2005, doi: 10.1145/1117454.1117456.

[107] A. Kumar, S. Kushwah, and A. K. Manjhvar, "A review on link prediction in social network," *Int. J. Grid Distrib. Comput.*, vol. 9, no. 2, pp. 43–50, 2016, doi: 10.14257/ijgdc.2016.9.2.05.

[108] S. Haghani and M. R. Keyvanpour, "A systemic analysis of link prediction in social network," *Artif. Intell. Rev.*, pp. 1–35, 2017, doi: 10.1007/s10462-017-9590-2.
[109] B. Chen, Y. Ding, and D. J. Wild, "Assessing drug target association using semantic linked data," *PLoS Comput. Biol.*, vol. 8, no. 7, 2012, Art. no. e1002574, doi: 10.1371/journal.pcbi.1002574.
[110] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Phys. Rep.*, vol. 659, pp. 1–44, Nov. 2016, doi: 10.1016/j.physrep.2016.09.002.
[111] M. Pourkazemi and M. R. Keyvanpour, "Community detection in social network by using a multi-objective evolutionary algorithm," *Intell. Data Anal.*, vol. 21, no. 2, pp. 385–409, 2017, doi: 10.3233/IDA-150429.
[112] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010, doi: 10.1016/j.physrep.2009.11.002.
[113] S. E. Schaeffer, "Graph clustering," *Comput. Sci. Rev.*, vol. 1, no. 1, pp. 27–64, Aug. 2007, doi: 10.1016/j.cosrev.2007.05.001.
[114] M. Pourkazemi and M. Keyvanpour, "A survey on community detection methods based on the nature of social networks," in *Proc. 3rd Int. Conf. Comput. Knowl. Eng.*, 2013, pp. 114–120, doi: 10.1109/ICCKE.2013.6682855.
[115] L. Udrescu, et al., "Clustering drug-drug interaction networks with energy model layouts: community analysis and drug repurposing," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 32745, doi: 10.1038/srep32745.
[116] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis Lectures Artif. Intell. Mach. Learn.*, vol. 3, no. 1, pp. 1–130, Jan. 2009, doi: 10.2200/S00196ED1V01Y200906AIM006.
[117] Y.-F. Huang, H.-Y. Yeh, and V.-W. Soo, "Inferring drug-disease associations from integration of chemical, genomic and phenotype data using network propagation," *BMC Med. Genomics*, vol. 6, no. Suppl 3, 2013, Art. no. S4, doi: 10.1186/1755-8794-6-S3-S4.
[118] G. Wu, J. Liu, and C. Wang, "Semi-supervised graph cut algorithm for drug repositioning by integrating drug, disease and genomic associations," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2016, pp. 223–228, doi: 10.1109/BIBM.2016.7822522.
[119] G. Wu, J. Liu, and C. Wang, "Predicting drug-disease interactions by semi-supervised graph cut algorithm and three-layer data integration," *BMC Med. Genomics*, vol. 10, no. Suppl 5, pp. 17–30, 2017, doi: 10.1186/s12920-017-0311-0.
[120] H. Te Yang, J. H. Ju, Y. T. Wong, I. Shmulevich, and J. H. Chiang, "Literature-based discovery of new candidates for drug repurposing," *Briefings Bioinf.*, vol. 18, no. 3, pp. 488–497, 2017, doi: 10.1093/bib/bbw030.
[121] A. Ezzat, M. Wu, X.-L. Li, and C.-K. Kwoh, "Computational prediction of drug–target interactions using chemogenomic approaches: An empirical survey," *Briefings Bioinf.*, 2018, Art. no. bby002, doi: 10.1093/bib/bby002.
[122] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces," *Bioinf.*, vol. 24, no. 13, pp. 232–240, 2008, doi: 10.1093/bioinformatics/btn162.
[123] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, and A. Zhavoronkov, "Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data," *Mol. Pharmaceutics*, vol. 13, no. 7, pp. 2524–2530, 2016, doi: 10.1021/acs.molpharmaceut.6b00248.
[124] N. C. Baker, S. Ekins, A. J. Williams, and A. Tropsha, "A bibliometric review of drug repurposing," *Drug Discovery Today*, vol. 23, no. 3, pp. 661–672, 2018, doi: 10.1016/j.drudis.2018.01.018.
[125] S. N. Deftereos, C. Andronis, E. J. Friedla, A. Persidis, and A. Persidis, "Drug repurposing and adverse event prediction using high-throughput literature analysis," *Wiley Interdisciplinary Rev. Syst. Biol. Med.*, vol. 3, no. 3, pp. 323–334, 2011, doi: 10.1002/wsbm.147.
[126] R. Leaman, R. I. Dogan, and Z. Lu, "DNorm: Disease name normalization with pairwise learning to rank," *Bioinf.*, vol. 29, no. 22, pp. 2909–2917, 2013, doi: 10.1093/bioinformatics/btt474.
[127] J. Li and Z. Lu, "Systematic identification of pharmacogenomics information from clinical trials," *J. Biomed. Inform.*, vol. 45, no. 5, pp. 870–878, 2012, doi: 10.1016/j.jbi.2012.04.005.
[128] R. Leaman, C.-H. Wei, and Z. Lu, "tmChem: A high performance approach for chemical named entity recognition and normalization," *J. Cheminformatics*, vol. 7, no. Suppl 1, 2015, Art. no. S3, doi: 10.1186/1758-2946-7-S1-S3.
[129] A. Lamurias and F. M. Couto, "Text Mining for Bioinformatics Using Biomedical Literature," *Encyclopedia of Bioinformatics and Computational Biology*, vol. 1, no. January, pp. 602–611, 2019, doi: 10.1016/B978-0-12-809633-8.20409-3.
[130] C. H. Wei, H. Y. Kao, and Z. Lu, "PubTator: A web-based text mining tool for assisting biocuration," *Nucl. Acids Res.*, vol. 41, no. Web Server issue, pp. 518–522, 2013, doi: 10.1093/nar/gkt441.
[131] M. Rastegar-Mojarad, R. K. Elayavilli, D. Li, R. Prasad, and H. Liu, "A new method for prioritizing drug repositioning candidates extracted by literature-based discovery," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, pp. 669–674, 2015, doi: 10.1109/BIBM.2015.7359766.
[132] F. Ghamami and M. Keyvanpour, "Why biomedical relation extraction is an open issue?," *ICIC Express Lett.*, vol. 9, no. 8, pp. 747–756, 2018, doi: 10.24507/icicelb.09.08.747.
[133] I. Mar Inez, F. Berzal, and J. Cubero, "A survey of link prediction in complex networks," *ACM Comput. Surv. Artic.*, vol. 49, no. 4, 2016, Art. no. 69, doi: 10.1145/3012704.
[134] Z. Yang, W. W. Cohen, and R. Salakhutdinov, "Revisiting semi-supervised learning with graph embeddings," in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn.*, 2016, vol. 48, pp. 40–48.
[135] A. S. Brown and C. J. Patel, "A review of validation strategies for computational drug repositioning," *Briefings Bioinf.*, vol. 19, no. 1, pp. 174–177, Jan. 2018, doi: 10.1093/bib/bbw110.
[136] R. Lichtenwalter and N. V. Chawla, "Link prediction: Fair and effective evaluation," *Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw . Anal. Mining*, pp. 376–383, 2012, doi: 10.1109/ASONAM.2012.68.
[137] Y. Yang, R. N. Lichtenwalter, and N. V. Chawla, "Evaluating link prediction methods," pp. 1–35, 2015, doi: 10.1007/s10115-014-0789-0.
[138] A. S. Brown and C. J. Patel, "Data descriptor:A standard database for drug repositioning," *Sci. Data*, vol. 4, no. 1, Apr. 2017, Art. no. 170029, doi: 10.1038/sdata.2017.29.
[139] K. Shameer, et al., "Systematic analyses of drugs and disease indications in RepurposeDB reveal pharmacological, biological and epidemiological factors influencing drug repositioning," *Briefings Bioinf.*, vol. 19, no. 4, pp. 656–678, Jul. 2018, doi: 10.1093/bib/bbw136.
[140] Z. Liu, A. Zhang, J. Ji, C. Liu, and X. Quan, "Survey: Functional module detection from protein-protein interaction networks," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 261–277, Feb. 2014, doi: 10.1109/TKDE.2012.225.
[141] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240, doi: 10.1145/1143844.1143874.
[142] T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *J. Biomed. Inf.*, vol. 38, no. 5, pp. 404–415, 2005, doi: 10.1016/j.jbi.2005.02.008.
[143] C. H. Nguyen and H. Mamitsuka, "Latent feature Kernels for link prediction on sparse graphs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 11, pp. 1793–1804, Nov. 2012, doi: 10.1109/TNNLS.2012.2215337.
[144] M. Kalra and N. Lal, "Data mining of heterogeneous data with research challenges," in *Proc. Symp. Colossal Data Anal. Netw.*, 2016, pp. 1–6, doi: 10.1109/CDAN.2016.7570899.
[145] N. Dogan and Z. Tanrikulu, "A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness," *Inf. Technol. Manage.*, vol. 14, no. 2, pp. 105–124, 2013, doi: 10.1007/s10799-012-0135-8.
[146] Y. Yao, Z. Xiao, B. Wang, B. Viswanath, H. Zheng, and B. Y. Zhao, "Complexity vs. performance," in *Proc. Internet Meas. Conf.*, 2017, vol. 1417, no. 119, pp. 384–397, doi: 10.1145/3131365.3131372.
[147] D. M. Dunlavy, T. G. Kolda, and E. Acar, "Temporal link prediction using matrix and tensor factorizations," *ACM Trans. Knowl. Discovery Data*, vol. 5, no. 2, pp. 1–27, Feb. 2011, doi: 10.1145/1921632.1921636.
[148] H. Litwin and K. J. Stoeckel, "Social network, activity participation, and cognition," *Res. Aging*, vol. 38, no. 1, pp. 76–97, Jan. 2016, doi: 10.1177/0164027515581422.
[149] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proc. IEEE*, vol. 104, no. 1, pp. 11–33, Jan. 2016, doi: 10.1109/JPROC.2015.2483592.
[150] A. Khaksari, and M. Keyvanpour, "TP-TA: A comparative analytical framework for trust prediction models in online social networks based on trust aspects," *Artif. Intell. Rev.*, 2017, doi: 10.1007/s10462-017-9583-1.
[151] W. Liu, J. Wang, and S.-F. Chang, "Robust and scalable graph-based semisupervised learning," in *Proc. IEEE*, vol. 100, no. 9, pp. 2624–2638, Sep. 2012, doi: 10.1109/JPROC.2012.2197809.
[152] A. Manconi, E. Vargiu, G. Armano, and L. Milanesi, "Literature retrieval and mining in bioinformatics: State of the art and challenges," *Adv. Bioinf.*, vol. 2012, pp. 1–10, 2012, doi: 10.1155/2012/573846.

[153] 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, Nov. 2013, Art. no. e78518, doi: 10.1371/journal.pone.0078518.
[154] H. Luo, et al., "Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm," *Bioinf.*, vol. 32, no. 17, pp. 2664–2671, Sep. 2016, doi: 10.1093/bioinformatics/btw228.
[155] X. Zeng, S. Zhu, X. Liu, Y. Zhou, R. Nussinov, and F. Cheng, "deepDR: A network-based deep learning approach to in silico drug repositioning," *Bioinf.*, 2019, Art. no. btz418, doi: 10.1093/bioinformatics/btz418.
[156] Z. Kuang, et al., "A machine-learning-based drug repurposing approach using baseline regularization," *Comput. Methods Drug Repurposing*, vol. 1903, pp. 255–267, 2019, doi: 10.1007/978-1-4939-8955-3_15.
[157] H. Xu, et al., "Validating drug repurposing signals using electronic health records: A case study of metformin associated with reduced cancer mortality," *J. Am. Med. Inf. Assoc.*, vol. 22, no. 1, pp. 179–191, 2015, doi: 10.1136/amiajnl-2014-002649.
[158] A. K. Menon and C. Elkan, "Link prediction via matrix factorization," in *Machine Learning and Knowledge Discovery in Databases*, Berlin, Germany: Springer, 2011, pp. 437–452, doi: 10.1007/978-3-642-23783-6_28.
[159] H. Luo, M. Li, S. Wang, Q. Liu, Y. Li, and J. Wang, "Computational drug repositioning using low-rank matrix approximation and randomized algorithms," *Bioinf.*, vol. 34, no. 11, pp. 1904–1912, Jun. 2018, doi: 10.1093/bioinformatics/bty013.
[160] B. Chen, F. Li, S. Chen, R. Hu, and L. Chen, "Link prediction based on non-negative matrix factorization," *PLoS One*, vol. 12, no. 8, pp. 1–18, Aug. 2017, doi: 10.1371/journal.pone.0182968.

Seyedeh Shaghayegh Sadeghi received the BS degree in computer science from Alzahra University, Iran, in 2015. She is currently working toward the master's degree in the Department of Computer Engineering and Data Mining Lab, Alzahra University, Tehran, Iran. Her main research interests include bioinformatics and data mining.

Mohammad Reza Keyvanpour received the BS degree in software engineering from the Iran University of Science & Technology, in 1997, and the MS and PhD degrees in software engineering from Tarbiat Modares University, Tehran, Iran, in 2000 and 2007, respectively. Currently, he is an associate professor with Alzahra University, Tehran, Iran. His current research interests include E-health and data mining.