OXFORD

Briefings in Bioinformatics, 2024, **25(3)**, bbae172

https://doi.org/10.1093/bib/bbae172

Problem Solving Protocol

A comparative benchmarking and evaluation framework for heterogeneous network-based drug repositioning methods

Yinghong Li †, Yinqi Yang†, Zhuohao Tong, Yu Wang, Qin Mi, Mingze Bai, Guizhao Liang, Bo Li and Kunxian Shu

Corresponding author: Yinghong Li, Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, No. 2, Chongwen Road, Nan'an District, Chongqing 400065, China. E-mail: liyinghong@cqupt.edu.cn; Bo Li, College of Life Sciences, Chongqing Normal University, 12 Tianchenlu, Shapingba District, Chongqing 400065, China. E-mail: libcell@cqnu.edu.cn; Kunxian Shu, Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, No. 2, Chongwen Road, Nan'an District, Chongqing 400065, China. E-mail: shukx@cqupt.edu.cn

†Yinghong Li and Yinqi Yang contributed equally to this work.

Abstract

Computational drug repositioning, which involves identifying new indications for existing drugs, is an increasingly attractive research area due to its advantages in reducing both overall cost and development time. As a result, a growing number of computational drug repositioning methods have emerged. Heterogeneous network-based drug repositioning methods have been shown to outperform other approaches. However, there is a dearth of systematic evaluation studies of these methods, encompassing performance, scalability and usability, as well as a standardized process for evaluating new methods. Additionally, previous studies have only compared several methods, with conflicting results. In this context, we conducted a systematic benchmarking study of 28 heterogeneous network-based drug repositioning methods on 11 existing datasets. We developed a comprehensive framework to evaluate their performance, scalability and usability. Our study revealed that methods such as HGIMC, ITRPCA and BNNR exhibit the best overall performance, as they rely on matrix completion or factorization. HINGRL, MLMC, ITRPCA and HGIMC demonstrate the best performance, while NMFDR, GROBMC and SCPMF display superior scalability. For usability, HGIMC, DRHGCN and BNNR are the top performers. Building on these findings, we developed an online tool called HN-DREP (http://hn-drep.lyhbio.com/) to facilitate researchers in viewing all the detailed evaluation results and selecting the appropriate method. HN-DREP also provides an external drug repositioning prediction service for a specific disease or drug by integrating predictions from all methods. Furthermore, we have released a Snakemake workflow named HN-DRES (https://github.com/lyhbio/HN-DRES) to facilitate benchmarking and support the extension of new methods into the field.

Yinghong Li is an associate professor at Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, China. His expertise lies in the fields of bioinformatics, drug discovery, and artificial intelligence.

Yinqi Yang is a master's degree candidate at the School of Bioinformatics, Chongqing University of Posts and Telecommunications, and a member of the Genomics and Intelligent Computing team. His current research interests include bioinformatics and computational biology.

Zhuohao Tong is a master's degree candidate at the School of Bioinformatics, Chongqing University of Posts and Telecommunications, with interests in indentification of prognostic and predictive biomarkers and in biomarker database construction research.

Yu Wang is a master's degree candidate at the School of Bioinformatics, Chongqing University of Posts and Telecommunications. His research interests include deep learning, natural language processing and bioinformatics.

Qin Mi is a master's degree candidate at the School of Bioinformatics, Chongqing University of Posts and Telecommunications, and a member of the Genomics and Intelligent Computing team. Her current research interests include bioinformatics and computational biology.

Mingze Bai is a professor at Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, China. His specialized research areas are bioinformatics, proteomics, and artificial intelligence.

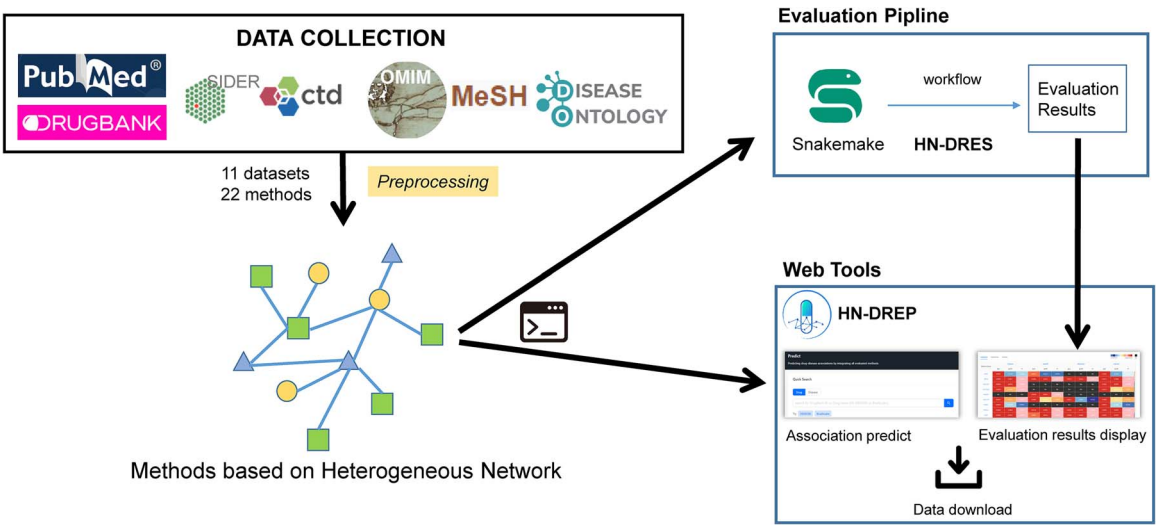
Guizhao Liang is a professor at Key Laboratory of Biorheological Science and Technology, Ministry of Education, Bioengineering College, Chongqing University, Chongqing, China. His current research interests are in bioinformatics, mathematical modeling, and molecular nutrition.

Bo Li is an associate professor at the College of Life Sciences, Chongqing Normal University, China. He earned his PhD from Chongqing University in China and currently focuses on the fields of bioinformatics, computational biology, and systems biology.

KunXian Shu is a professor at Chongqing Key Laboratory of Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, China. His specialized research areas are bioinformatics, systems biology, and artificial intelligence.

Received: December 10, 2023. **Revised:** February 25, 2024. **Accepted:** April 2, 2024.

Graphical Abstract



Keywords: drug repositioning; heterogeneous networks; method evaluation; online tools; evaluation workflow

INTRODUCTION

The traditional process of designing and discovering new drugs is time-consuming, costly and risky [1]. In light of these challenges, computational drug repositioning (also known as drug repurposing), which aims to find new indications for approved and clinical drugs, has emerged as an alternative to traditional drug discovery [2]. Because these drugs have safety, efficacy and tolerability data from preliminary testing and clinical trials, drug repositioning can facilitate drug discovery and reduce overall development costs [3]. Therefore, drug repositioning is an effective strategy for drug discovery and is increasingly becoming an attractive research topic [4].

With the development of multiomics data, high-throughput sequencing technologies and continuously updated databases, an abundance of computational methods have been proposed to predict potential drug–disease associations for drug repositioning [4–8]. Heterogeneous network-based approaches, which utilize the relationships among biomedical entities to construct heterogeneous networks with the ability to integrate multiple data sources, are widely used in drug repositioning research [4, 9–11]. Moreover, these approaches have been shown to outperform other methods by capturing similar information in different biological networks as drug and disease features to improve the accuracy of drug repositioning and have thus become the predominant and widely embraced choice in this field [5, 12, 13]. Therefore, in this paper, we focus on recent heterogeneous network-based drug repositioning methods.

Current heterogeneous network-based methods can be roughly divided into three categories based on their algorithm: machine learning-based methods, network propagation-based methods and matrix completion or factorization-based methods [14, 15]. Furthermore, methods derived from heterogeneous networks can be broadly categorized based on their underlying network structures into bipartite networks-based methods, tripartite networks-based methods and other complex networks-based methods. For instance, the ANMF [16] devised by Yang et al. is anchored in the bipartite networks-based method (drug-disease); the HINGRL [12] developed by Zhao et al. is an example of tripartite networks-based method; while Daniel et al.' Drug2ways [17], leveraging knowledge graphs, is considered to be other complex networks-based methods. Considering the preponderance of heterogeneous network-based methods are either bipartite or tripartite, our investigation predominantly explores these two categories.

These computational methods use the principle of 'guilt-by-association' to discover new indications of existing drugs [18, 19], which assumes that similar drugs are associated with similar diseases and vice versa. While all network-based drug repositioning methods share a common goal of finding new indications for existing drugs, they differ in terms of underlying algorithms and needed input data. With the increasing number of heterogeneous network-based drug repositioning methods being developed, researchers new to the field or wishing to analyze new datasets are faced with a multitude of method choices, and it is not clear which method will best address their problems.

Given the diversity of heterogeneous network-based drug repositioning methods, it is important to quantitatively assess their performance, scalability and usability. However, the existing methods have only been compared with several methods in some studies, and there are conflicting results in different studies. For example, while Xie et al. [14] contend that the BNNR method outperforms HGIMC, other researchers, such as Yang et al. [15] and Yan et al. [20], argue that HGIMC surpasses BNNR in terms of performance. A recent review study evaluated the AUC and AUPR values of 11 drug repositioning methods based on heterogeneous networks on two datasets [10]. However, the paper was limited in its scope, as it only evaluated the predictive performance of these methods. The scalability and usability of these methods were not compared or analyzed. To the best of our knowledge, a comprehensive comparison of heterogeneous network-based drug repositioning methods across a large number of different datasets is still lacking, and the criteria for the evaluation and comparison of methods vary. More importantly, the strengths and weaknesses of existing methods must be assessed to guide the

Table 1: Drug repositioning methods based on heterogeneous networks in this study

Method	Platform	Networks	Algorithms	Category	Reference
ANMF	Python	Bipartite network (drug–disease)	Autoencoder	Machine Learning	[16]
BNNR	Matlab	Bipartite network (drug–disease)	Nuclear norm regularization, ADMM	Matrix Completion	[5]
DDAGDL	Python	Tripartite network (drug-disease-protein)	Geometric deep learning, XGBoost, autoencoder	Machine Learning	[21]
DDAPRED	Python	Bipartite network (drug–disease)	Logistic matrix factorization, similar network fusion	Matrix Factorization	[22]
DDA-SKF	Matlab	Bipartite network (drug–disease)	Similarity kernel fusion, Laplacian regularized least squares	Machine Learning	[23]
deepDR	Python	–	MDA, cVAE	Machine Learning	[24]
DRAGNN	Python	Bipartite network (drug-disease)	GNN, attention, MLP	Machine Learning	[25]
DRHGCN	Python	Bipartite network (drug–disease)	Graph convolutional network	Machine Learning	[6]
DRIMC	R	Bipartite network (drug–disease)	Logistic matrix factorization	Matrix Completion	[26]
DRPADC	Matlab	Bipartite network (drug–disease)	WKNKN, CKA-MKL	Matrix Completion	[14]
DRRS	Matlab	Bipartite network (drug–disease)	SVT, nuclear norm minimization	Matrix Completion	[13]
DRWBNCF	Python	Bipartite network (drug-disease)	MLP, weighted bilinear aggregator	Machine Learning	[27]
GROBMC	Matlab	Bipartite network (drug–disease)	Laplacian graph regularization, nuclear norm minimization, PPXA	Matrix Completion	[28]
HGIMC	Matlab	Bipartite network (drug–disease)	HGBI, bounded matrix completion, Gaussian radial basis, ADMM	Matrix Completion	[15]
HINGRL	Python	Tripartite network (drug–disease-protein)	Random walk, autoencoder	Network Propagation	[12]
HNRD	Python	Bipartite network (drug–disease)	Neighborhood information aggregation, Neural network	Machine Learning	[29]
iDrug	Matlab	Tripartite network (drug–disease-target)	Cross-network embedding, multiplicative update minimization	Network Propagation	[30]
ITRPCA	Matlab	Bipartite network (drug-disease)	WKNN, TRPCA	Matrix Completion	[31]
LAGCN	Python	Bipartite network (drug–disease)	Graph convolutional network	Machine Learning	[32]
MBiRW	Matlab	Bipartite network (drug–disease)	Bi-random walk	Network Propagation	[33]
MLMC	Matlab	Bipartite network (drug–disease)	Laplacian graph regularization, ADMM	Matrix Completion	[20]
MSBMF	Matlab	Bipartite network (drug–disease)	Bilinear matrix factorization, ADMM	Matrix Factorization	[11]
NMFDR	Matlab	Bipartite network (drug–disease)	Non-negative matrix factorization, Similarity Network Fusion	Matrix Factorization	[34]
OMC	Matlab	Bipartite network (drug–disease)	Nuclear norm minimization, ADMM,KNN	Matrix Completion	[35]
SCMFDD	Matlab	Bipartite network (drug–disease)	Similarity constrained matrix factorization	Matrix Factorization	[36]
SCPMF	Matlab	Bipartite network (drug–disease)	Similarity constrained probabilistic matrix factorization	Matrix Factorization	[37]
VDA-GKSBMF	Matlab	Bipartite network (drug-disease)	Gaussian kernel similarity bilinear matrix factorization, ADMM	Matrix Factorization	[38]
WRMF	Matlab	Bipartite network (drug-disease)	Similarity constrained weight regularization matrix factorization	Matrix Factorization	[39]

development of new methods that can improve upon the current state-of-the-art.

Here, we present a comprehensive evaluation of the performance, scalability and usability of 28 heterogeneous network-based drug repositioning methods using 11 datasets. We also developed a standardized evaluation process, HN-DRES, for new methods. Additionally, we created an interactive website, HN-DREP, to facilitate user access to evaluation results, selection of appropriate drug repositioning methods and datasets, and drug repositioning for drugs or diseases of interest. Our evaluation provides valuable insights for the development of new methods and promotes the advancement of drug repositioning research.

Table 2: Datasets utilized in this study

Datasets	No. of drugs	No. of diseases	No. of associations	Disease ID	Reference
Fdataset	593	313	1933	OMIM	[43]
Cdataset	663	409	2352	OMIM	[33]
DNdataset	1490	4516	1008	–	[44]
iDrug	1321	3966	111 481	OMIM	[30]
Ydataset	1478	655	8448	OMIM	[11]
LRSSL	763	681	3051	MeSH	[45]
LAGCN	269	598	18 416	MeSH	[36]
SCMFDD_L	1323	2834	49 217	MeSH	[36]
deepDR	1519	1229	6677	MedGen	[24]
HDVD	219	34	455	–	[37]
TLHGBI	1409	5080	1461	–	[46]

- indicates that the disease ID in this dataset is missing and is replaced by the disease name.

RESULTS

Overview of the involved methods and datasets

To identify drug repositioning methods based on heterogeneous networks, we conducted a systematic literature review of PubMed articles published before 30 December 2023, using the search terms '(drug repositioning method[Title/Abstract]) AND (heterogeneous network[Title/Abstract])', '(drug repurposing method [Title/Abstract]) AND (heterogeneous network[Title/Abstract])' and '(drug reprofiling method[Title/Abstract]) AND (heterogeneous network[Title/Abstract])'. We identified 170 methods (Supplementary Data 1); however, upon review, we found that most were not feasible for practical use due to one or more of the following criteria: (i) unavailable or unusable code; (ii) predictions not limited to drugs and diseases; (iii) missing material (code or data) or (iv) additional inputs needed during algorithm execution. Ultimately, our evaluation included 28 methods (Table 1, Supplementary Material—Supplementary Note 1).

To comprehensively evaluate the performance of heterogeneous network-based drug repositioning methods, we collected 11 benchmark datasets, which include all datasets used in existing drug repositioning studies (Table 2, Supplementary Material—Supplementary Note 2). Fdataset and Cdataset are two widely adopted gold standard datasets in method comparison. Specifically, the drugs in most of these datasets are from the DrugBank database [40], while the diseases are from three independent and incompatible databases: the Online Mendelian Inheritance in Man (OMIM) database [41], Comparative Toxicogenomics Database (CTD) database [42] and MeSH (medical subject headings vocabulary).

The overall benchmark framework

We conducted a comprehensive evaluation of the performance, scalability and usability of 28 heterogeneous network-based drug repositioning methods across 11 existing datasets. Specifically, our evaluation strategy comprised three components (Figure 1): (i) Performance evaluation: we performed 10-fold cross-validation on the results of each method on each dataset and evaluated their performance using multiple metrics, including precision, recall, F1 score and area under the ROC curve (AUC); (ii) Scalability evaluation: we calculated the running time and peak memory usage for each method to predict drug-disease association results on each dataset; (3) Usability evaluation: we quantified the usability of each method using a transparent scoring scheme that considered factors such as documentation, ease of use and flexibility.

Additionally, we established a standardized workflow called HN-DRES (Heterogeneous Network-based Drug Repositioning method Evaluation Snakemake workflow) to simplify the evaluation task and assess its output.

Overall performance

We categorized the evaluated methods into three main groups and applied 10-fold cross-validation to each method on each dataset to calculate their respective performance metrics. Additionally, we assessed the scalability and usability of each method to comprehensively evaluate and compare existing drug repositioning approaches.

Our findings revealed significant variation in the performance of these methods across different datasets, emphasizing the lack of a one-size-fits-all approach. Moreover, we did not observe a clear correlation among several evaluation criteria. Nevertheless, aggregating the evaluation results from the three aspects, our overall score (Figure 2B) highlighted that matrix completion or factorization methods, such as HGIMC, ITRPCA and BNNR, generally demonstrated strong performance across the board. However, it is worth noting that these methods excelled in different aspects according to various evaluation criteria. For instance, HINGRL, MLMC and HGIMC displayed higher performance, while NMFDR, GROBMC and SCPMF exhibited superior scalability. Meanwhile, HGIMC, DRHGCN and BNNR stood out for their usability (Supplementary Data 2). We discuss the specifics of each evaluation aspect in detail below.

Furthermore, we observed that the overall performance of drug repositioning methods based on heterogeneous networks is not correlated with the network type (bipartite or tripartite). There seems to be no straightforward association between the effectiveness of the methods and the specific types of networks. For example, although the well-performing HGIMC, ITRPCA and BNNR methods are bipartite networks, the poorly performing LAGCN and ANMF methods are also bipartite networks.

Method performance

To assess the predictive performance of the methods, we computed several widely used metrics in the field of drug repositioning and aggregated their scores on different aspects. Three distinct metrics were employed to evaluate method performance, each offering a unique perspective: AUC, AUPR and F1 score (Figure 3C). Based on the overall scores of the three selected metrics, HINGRL (0.892), MLMC (0.876), ITRPCA (0.863) and HGIMC (0.863) exhibited the best performance. Conversely, DRIMC (0.416), ANMF

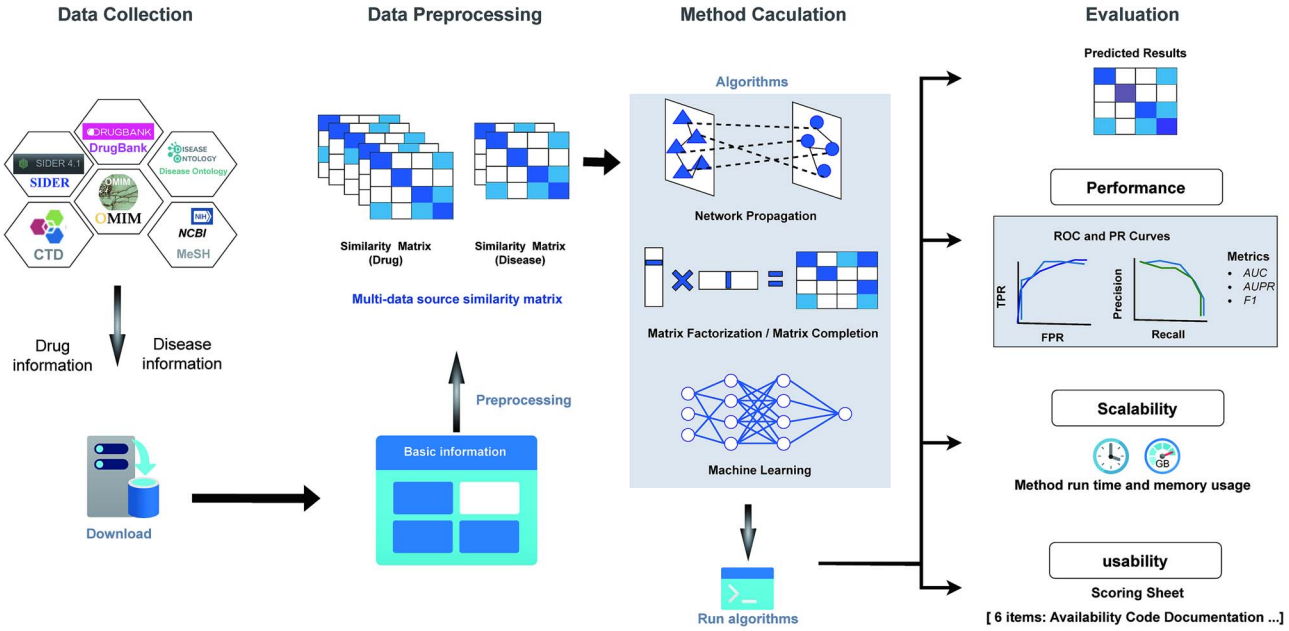


Figure 1. The overall benchmark framework.

(0.416), SCMFDD (0.39) and LAGCN (0.38) received comparatively lower scores (Figure 3B, Supplementary Data 2). We observed that matrix completion or factorization methods performed better overall. However, the performance of different methods varies greatly across datasets, and some methods may achieve unexpected results on a particular dataset. Therefore, users should try different methods on their data, as no single method emerged as universally superior across all datasets.

Among all methods, GROBMC exhibited the highest overall AUC score of 0.969, followed by MLMC (0.965), DRIMC (0.959), ITRPCA (0.947) and HGIMC (0.945). GROBMC also achieved the highest AUC scores on five of the top 10 datasets, including iDrug, SCMFDD-L, Ydataset, Cdataset and HDVD. Notably, SCMFDD achieved the highest AUC score of 0.991 on the TLHGBI dataset and 0.986 on the DNdataset; however, its overall score was lower due to less favorable performance on smaller datasets, such as an AUC score of 0.776 on the HDVD dataset. LAGCN had the lowest overall AUC score of 0.836. On the widely used gold standard dataset Fdataset, GROBMC, MLMC and DRIMC demonstrated top-tier performance with AUC scores of 0.977, 0.959 and 0.957, respectively (Supplementary Data 2, Supplementary Data 3).

After a comprehensive analysis of the AUC metric obtained by the method on all datasets, GROBMC demonstrated the highest performance with an AUC score of 0.969. Impressively, five out of the top 10 highest AUC scores were attributed to GROBMC on the iDrug, SCMFDD-L, Ydataset, Cdataset and HDVD datasets. Additionally, MLMC (AUC: 0.965), DRIMC (AUC: 0.959), ITRPCA (0.947) and HGIMC (AUC: 0.945) displayed strong AUC scores, showcasing their predictive capabilities. Notably, SCMFDD achieved the highest AUC of 0.991 on the TLHGBI dataset and an impressive AUC of 0.986 on the DNdataset; however, its overall score was not as high due to less favorable performance on smaller datasets, such as an AUC of 0.776 on the HDVD dataset. The LAGCN method attained the lowest overall AUC value at 0.836 when compared with other methods. On the widely used gold standard dataset Fdataset, GROBMC (AUC: 0.977), MLMC (AUC: 0.959) and DRIMC (AUC: 0.957) demonstrated top-tier performance (Supplementary Data 2, Supplementary Data 3).

Among all methods evaluated on all datasets, MLMC emerged as the top performer overall with an AUPR score of 0.97, followed closely by HGIMC (AUPR: 0.957), ITRPCA (0.956), DDA-SKF (AUPR: 0.95), OMC (AUPR: 0.9444), deepDR (AUPR: 0.941), DRRS (AUPR: 0.94), BNNR (AUPR: 0.939), VDA-GKSBMF (0.937), NMFDR (AUPR: 0.934) and DRPADC (AUPR: 0.928). Conversely, LAGCN, DRWBNCF, SCMFDD and ANMF exhibited the weakest performance, with AUPR scores of 0.26, 0.253, 0.182 and 0.145, respectively. Notably, the DRPADC method achieved the highest AUPR score of 0.991 on the TLHGBI dataset, accounting for 3 of the top 10 AUPR scores. Moreover, 8 of the top 10 AUPR scores were observed on the larger datasets TLHGBI, DNdataset and iDrug. On the widely used gold standard dataset Fdataset, MLMC (AUPR: 0.965), ITRPCA (AUPR: 0.963), OMC (AUPR: 0.956) and HINGRL (AUPR: 0.9515) outperformed the other methods (Supplementary Data 2, Supplementary Data 3).

HINGRL achieved the best overall F1 score of 0.847, followed by MLMC (0.692), ITRPCA (0.687), HGIMC (0.686) and DRRS (0.683). These methods outperformed the others with commendable results. Notably, HINGRL achieved an impressive F1 score of 0.879 on the gold standard dataset. Conversely, LAGCN exhibited the weakest performance across multiple datasets, including deepDR, TLHGBI, and DNdataset. DRPADC excelled in F1 performance on the TLHGBI, iDrug and Ydataset, but its overall F1 score was mid-range. DRIMC exhibited the lowest overall F1 score, with a particularly poor F1 score of 0.01 on the gold standard dataset Fdataset (Supplementary Data 2, Supplementary Data 3).

Method scalability

To evaluate the scalability of the methods, we sequentially executed each method on a range of existing datasets from small to large, measuring their runtime and peak memory consumption. Our experiments revealed that NMFDR, GROBMC, SCPMF, WRMF and iDrug exhibited superior overall performance. Notably, NMFDR emerged as the fastest and most memory-efficient method, while LAGCN and ANMF exhibited the worst overall scalability. Importantly, most methods demonstrated commendable scalability performance. LAGCN, ANMF and

Figure 2. Characteristics and overall evaluation results of the 28 methods evaluated in this study. (**A**) We describe the methods according to their classification, algorithm and programming language. (**B**) We integrated the scores of the three indicators of performance, scalability and usability to conduct a comprehensive overall evaluation.

DRHGCN were more sensitive to dataset size in terms of memory consumption than other methods. Additionally, the runtimes of LAGCN, ANMF, SCMFDD, HNRD, DRHGCN and DRRS were more sensitive to dataset size (Figure 4B, Supplementary Data 3).

Specifically, NMFDR exhibited the most efficient runtime performance, completing the Fdataset in just 10 s and the larger iDrug dataset (1321∗3966) in 52 s, comfortably remaining within the 1-min threshold. Several other methods, including GROBMC, SCPMF, WRMF, MSBMF, HGIMC and iDrug, also displayed favorable runtime performance. Conversely, LAGCN, HNRD, DRAGNN and DRRS were the most time-consuming methods, with LAGCN performing the worst, requiring over 3 h to complete the TLHGBI dataset (1409∗5080). Notably, the runtimes of LAGCN, ANMF, SCMFDD, HNRD, DRHGCN and DRRS are more time-consuming and sensitive to dataset size than those of the other methods (Figure 4B, Supplementary Data 3).

Upon analysis of memory consumption, DDAPRED demonstrated the lowest memory overhead, performing exceptionally well with memory consumption not exceeding 1 GB across the Fdataset, Cdataset and LRSSL datasets. DRRS, SCPMF, WRMF, NMFDR, MBiRW, DRPADC, VDA-GKSBMF and iDrug also exhibited superior memory performance relative to other methods. Overall, the memory consumption of most methods remained within reasonable limits, without imposing excessive memory usage. However, LAGCN and ANMF were notable exceptions, displaying relatively high memory consumption that increased exponentially with dataset size. For instance, LAGCN's memory usage exceeded 100 GB on the iDrug, DNdataset and TLHGBI datasets, while ANMF's memory consumption on the SCMFDD-L (1323∗2834) dataset surpassed 180 GB (Figure 4B, Supplementary Data 3). Notably, our findings indicate that LAGCN, ANMF and DRHGCN are more sensitive to dataset size, resulting in increased memory usage. This is an important consideration for users when choosing a method to use, especially for large datasets.

Method usability

To assess the usability of the methods, we conducted a comprehensive evaluation using a transparent checklist that considered software accessibility, code quality, documentation, error

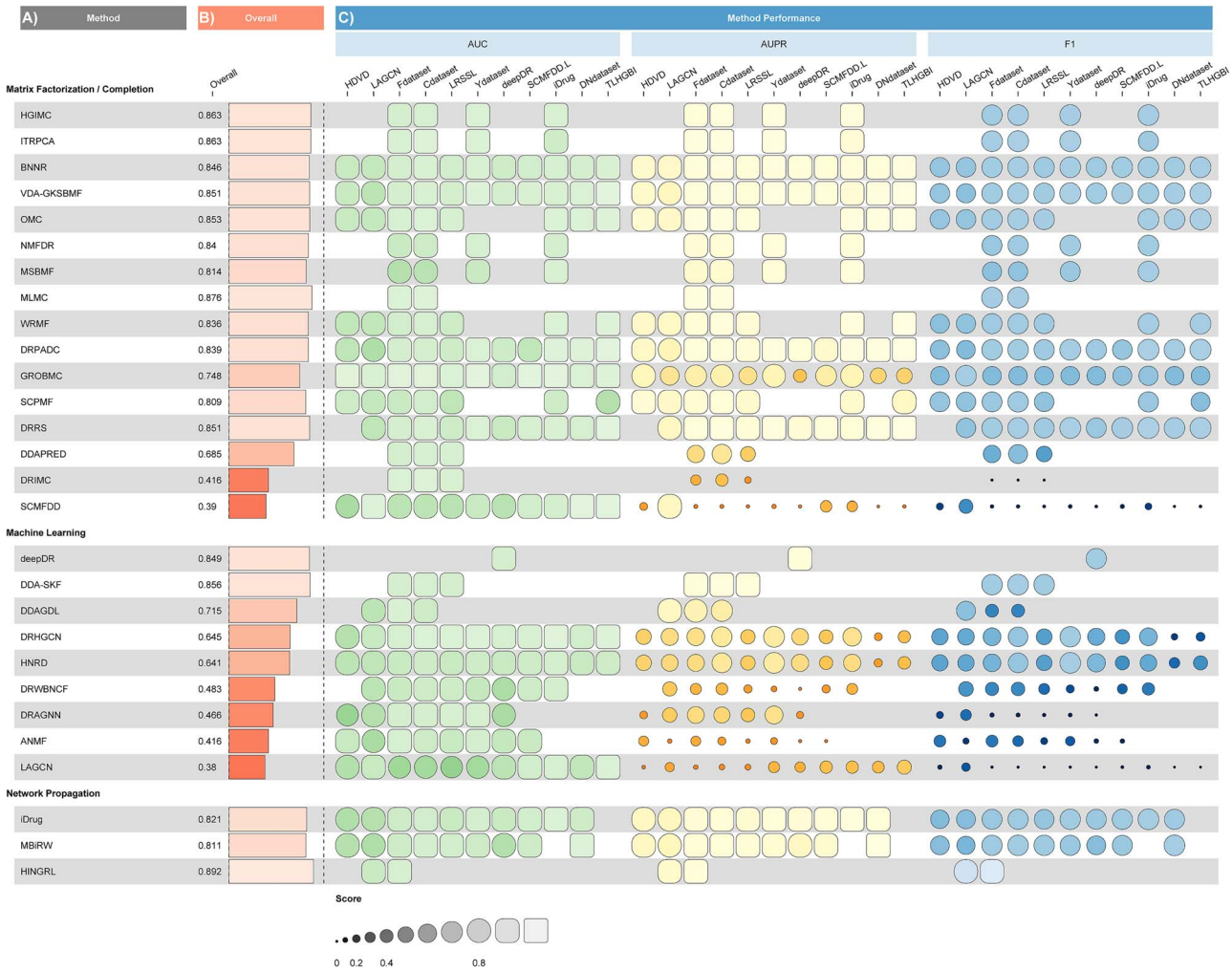


Figure 3. Performance results of methods on all datasets. (**A**) The names of the methods, sorted according to Figure 2. (**B**) Overall performance score of the method. (**C**) AUC, AUPR and F1 scores of each method on the dataset.

rates and other relevant factors (Figure 4C). Our findings revealed that most drug repositioning methods met the basic criteria, including accessibility, code availability and basic code quality (Supplementary Data 2).

However, we identified significant shortcomings in the documentation of method usage and the needed dependencies of the method execution environment for most methods. In many cases, these methods failed to specify the environment needed for their execution and lacked comprehensive documentation explaining their usage, which could be detrimental to users. Notably, our results indicate that only HGIMC and DRHGCN outperformed other methods in these aspects.

Furthermore, several methods, including ANMF, DRAGNN, DRWBNCF, MBiRW, SCPMF, WRMF, iDrug, OMC, DRRS and MLMC, exhibited execution errors such as out-of-memory errors, occurrences of null values and unsupported data formats. For example, the ANMF method encountered out-of-memory errors or missing values when processing large datasets such as iDrug, DNdataset and TLHGBI. Similarly, the SCPMF and WRMF methods may trigger out-of-memory errors or missing value issues with certain datasets. Additionally, the DRRS method encountered unspecified errors when applied to the HDVD dataset.

Overall, our findings highlight the importance of considering method quality and usability in the field of drug repositioning. While method availability does not directly correlate with method performance, methods with poor usability can be difficult and time-consuming to use, which can impede research progress.

HN-DREP web server tutorial

To facilitate access to the results of our study and to assist researchers in selecting the most appropriate methods for their specific needs, as well as in making drug repositioning predictions for the drugs or diseases of interest, we have developed an online tool, HN-DREP, which provides users with a free online service. Below is a brief overview of how to use HN-DREP.

The home page provides an overview of HN-DREP, its framework and statistics (Figure 5A). The Browse page lists the methods and datasets used in the study, along with their basic information (Figure 5B). The evaluation page presents the results of the method evaluation, including performance, scalability and usability (Figure 5C). Clicking on a method will direct users to its details page, which includes basic information, specific evaluation results on a particular dataset and the method's prediction

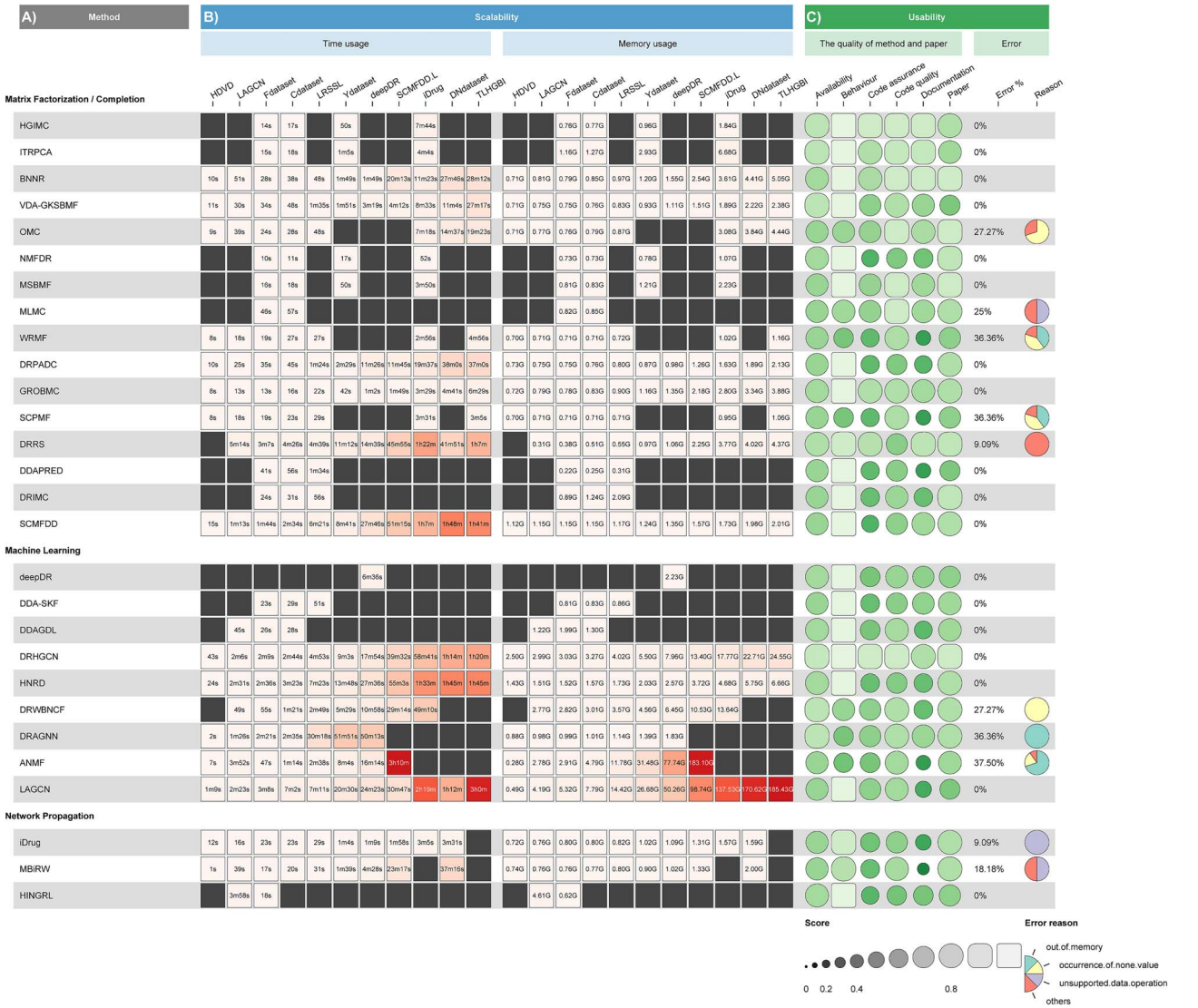


Figure 4. Method scalability and usability. (**A**) The names of the methods, sorted according to Figure 2. (**B**) The runtime and peak memory usage of each method on each dataset. (**C**) Usability of each method, including the reasons for errors and the error rates.

results. The Prediction page allows users to query for drug or disease predictions, and its Prediction Details page provides basic information about the selected entry, including the predicted drug or disease associated with it (Figure 5D). Finally, the Download page allows users to download the dataset, the drug-disease information it contains and the prediction results of the methods in the dataset (Figure 5E).

DISCUSSION

Summary of the study

In this investigation, we collated and screened 28 drug repositioning methods, as well as 11 datasets that have been utilized in extant drug repositioning studies. On this foundation, we conducted a comprehensive and systematic benchmark assessment of these drug repositioning methods. In addition to evaluating the methods' performance, we also assessed the scalability and usability of each method. Based on the results of our evaluation, in this work, we provide a stage-by-stage overview of the existing drug repositioning methods. We highlight methods demonstrating superior performance and underscore certain limitations in the extant drug repositioning approaches. These findings can offer valuable insights and guidance to researchers and developers working in the field of drug repositioning.

Explanations for the method's superior performance and some of our research findings

Our investigation reveals that methods in the category of matrix completion or factorization demonstrate noteworthy overall performance superiority; we attribute this to the limitations of traditional machine learning methods, which rely heavily on labeled samples within datasets [12, 16, 32]. In practical applications, acquiring stable sample data is often challenging, constraining the effectiveness and ability of these methods. Additionally, traditional machine learning methods are highly dependent on input data and feature extraction, making them less practical for real-world applications [14]. During the network propagation process in network propagation-based methods, information resources tend to favor edges with higher weights, which deprives nodes lacking associated information of resources for extended periods, resulting in the 'cold-start problem' [14]. This issue can affect the accuracy of the prediction results.

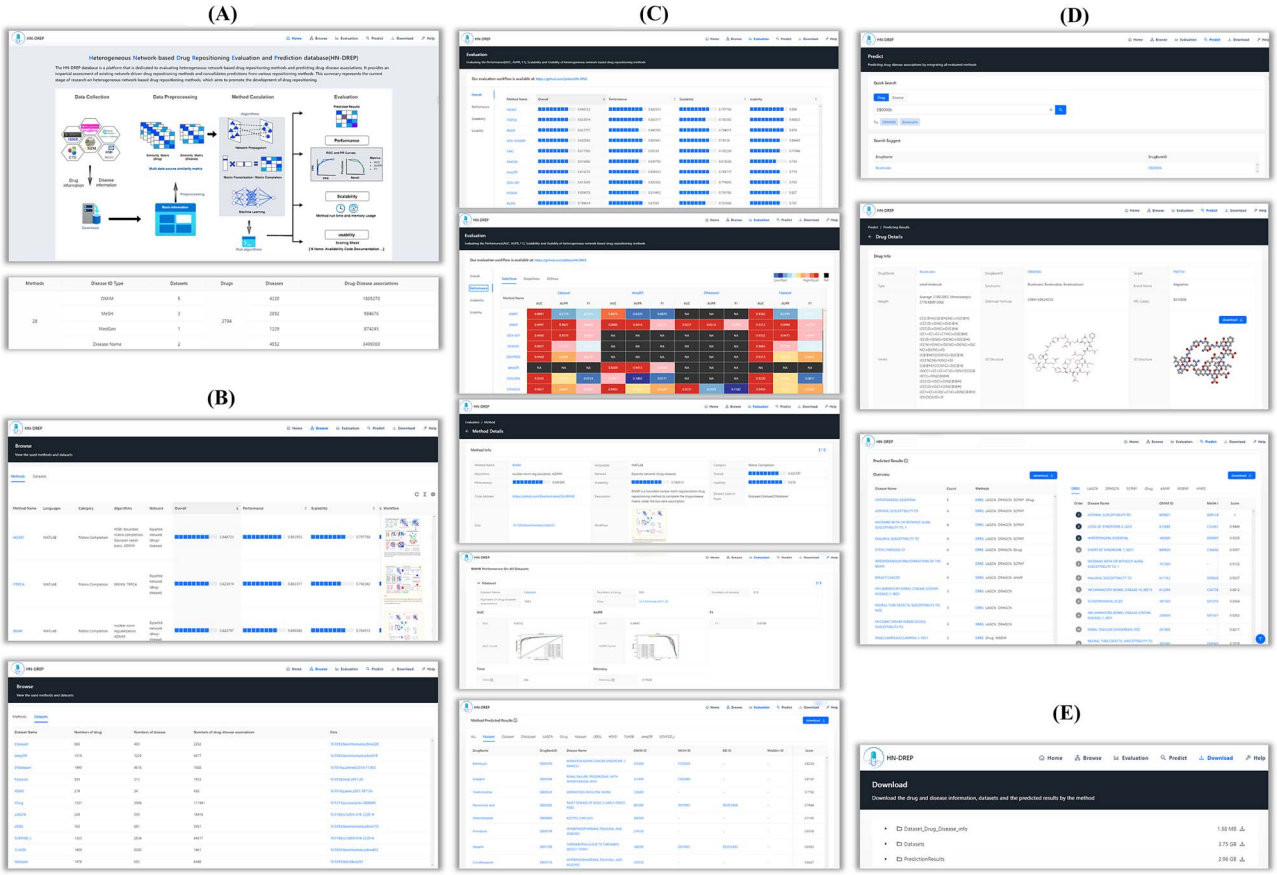


Figure 5. The interfaces of the web server: (**A**) the homepage, (**B**) methods and datasets, (**C**) evaluation page and its detail page, (**D**) prediction page and its detail page and (**E**) download page.

In contrast, matrix completion or matrix factorization methods use 'submatrix simulation' techniques, which are more flexible in integrating a priori information and do not rely heavily on predefined labels or negative samples. Instead, these methods extract implicit patterns from existing data matrices, capture the original matrix information through submatrices and generate low-rank simulation matrices to fill in the missing portions of the original association matrices [14, 34]. This approach does not require prior knowledge of extensive association information for predictions and has the advantages of adapting to sparse data, adapting to heterogeneous data, and scalability. Moreover, compared with other methods, matrix completion or matrix factorization methods consider all the main eigenvalues of the adjacency matrix and its associated eigenvectors [13], which reduces the redundancy of the model and enables mining of the association features between multiple similarity and association matrices [20, 32].

Our investigation revealed an unexpected absence of correlation between the efficacy of methods and the nature of the networks employed. Contrary to the conventional wisdom that tripartite networks, presumed to encapsulate more biologically significant information than bipartite networks, would facilitate more accurate predictions. However, our evaluation demonstrates that the distinction in performance between methods based on bipartite and tripartite networks is negligible.

Nevertheless, it is noteworthy that the efficacy of a method varies significantly across diverse datasets, underscoring the absence of a one-size-fits-all approach. Network propagation-based methods are advantageous in terms of computational efficiency [15]. Contrary to prior research [10], our study suggests that machine learning-based drug repositioning methods may consume more time and memory resources than methods such as matrix completion or factorization. The reproduction of drug repositioning methods based on heterogeneous networks frequently faces obstacles, including the absence of necessary documentation for software installation or setting up the execution environment, missing code or input data, non-operational code, and the requirement for additional inputs beyond heterogeneous network data during the method's execution process. It might be possible to address these challenges by creating a Docker image that ensures the method is operational. This strategy provides a straightforward way for editors or reviewers to determine the reproducibility of the method. A substantial proportion of methods lacked comprehensive documentation, failing to provide users with a detailed explanation of the method or clear instructions for establishing the necessary execution environment and reproducing the methods. Moreover, in many cases, researchers have unreasonably compared the performance of their published methods with the performance of other methods, selectively choosing the metrics or datasets in which their methods performed best, resulting in the biased outcome that their own methods tend to be superior [47].

Notably, the lack of gold standard datasets in drug repositioning and the translation of theoretical computational models into practice remain significant challenges [30, 48]. Although Gottlieb et al. [43] attempted to provide such datasets for practical use, they are outdated and incomplete. The application of standardized evaluation metrics and datasets is critical in this rapidly

evolving field of research, and new efforts are needed in this area [4].

Guidance on method selection and our offerings

Based on our assessment, we recommend that users carefully consider the specific characteristics of their dataset and research requirements when selecting a method. The optimal choice may not necessarily be the method with the highest performance but rather one that strikes a balance between performance, scalability and usability. To facilitate access to our collected methods, datasets and evaluation results, we developed an online platform, http://hn-drep.lyhbio.com/), which reduces the data collection burden on researchers and enables method selection based on individual user needs. Additionally, we offer drug repositioning prediction services to the public. We hope that our work will aid researchers and accelerate the advancement of drug repositioning research.

The shortcomings of our research

Despite providing a systematic evaluation of extant drug repositioning methods, our study has some limitations. For example, it did not encompass drug target association prediction methods, and the evaluation metrics focused primarily on AUC, AUPR and F1 scores, potentially overlooking other relevant performance metrics. Furthermore, the assessment of method time and memory consumption on different datasets may introduce noise, which could limit the accuracy of measuring method portability and potentially introduce bias.

Challenges and suggestions in drug repositioning

In view of the burgeoning research interest in drug repositioning methods, our study highlights several key challenges:

(1) In focusing on performance enhancement, developers should also prioritize the scalability, usability and documentation of their methods, alongside providing user-friendly execution environment tutorials, as these aspects collectively contribute significantly to the quality of publications.

(2) New method development should involve rigorous and comprehensive comparisons with existing high-performing methods, using diverse real datasets to advance the field and enhance method applicability.

(3) Leveraging multiomics data, high-throughput technologies and up-to-date databases is essential for developing new methods that can harness the wealth of available information to improve the accuracy and reliability of drug-disease associations in computational drug repositioning.

(4) The small size of the dataset limits its ability to discriminate between the different drug repositioning methods based on heterogeneous networks.

MATERIALS AND METHODS

Preprocessing

Drug similarity metrics were calculated based on chemical structures, ATC codes, side effects, drug–drug interactions and targets. Information on chemical structures, ATC codes, drug–drug interactions and targets was extracted from DrugBank, while side effect data were extracted from the SIDER [49] database. Drug chemical structure similarity was calculated using the R package RCDK [50], relying on SMILES files. ATC code similarity was calculated using the inverse document frequency and cosine similarity methods introduced by Kastrin et al. [51]. The similarity of side effects, drug–drug interactions and targets was computed using the Jaccard similarity coefficient [52].

For diseases, the disease phenotype similarity matrix was downloaded from the MimMiner database [53]. Disease ontology data were sourced from the Disease Ontology database [54] and processed using the R package DOSE [55] to generate the disease ontology similarity matrix.

Evaluation framework

All aforementioned methods are evaluated using a common evaluation pipeline. The importance of using multiple performance metrics to compare models or methods has been repeatedly emphasized and acknowledged [47, 56]. Therefore, to systematically evaluate the performance of drug repositioning methods, we evaluate each method using the following criteria (Supplementary Material—Supplementary Note 3).

The area under the receiver operating characteristic curve (AUC), the area under the precision–recall curve (AUPR) and the F1 score are widely used in bioinformatics research to evaluate the overall performance of drug repositioning methods [4]. We performed 10-fold cross-validation to evaluate the performance of each method using the Python package scikit-learn to calculate AUC, AUPR and F1 values.

For each method, a prediction score matrix is generated upon completion of the method's execution. Subsequently, a confusion matrix can be derived from this score matrix and the corresponding labels. For each specific ranking threshold, we calculate the values of true positive (TP), false-negative (FN), false-positive (FP) and true negative (TN). TP and TN indicate the correct identification of positive and negative samples, respectively, while FP and FN represent the incorrect identification of positive and negative samples. By varying the ranking threshold, the true positive rate (TPR), false-positive rate (FPR), precision, and recall can be calculated to construct the ROC curve and the precision-recall curve for visual comparison of method performance [57]. Ultimately, AUC, AUPR and F1 can be calculated and used to evaluate the overall performance of the drug repositioning method.

$$TPR\ (or\ Recall) = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

To assess the scalability of each method, we measured the peak memory usage and runtime consumption on each dataset using the '/usr/bin/time' command on Linux and Snakemake's benchmark directive. It is important to note that memory and time consumption estimates can vary considerably due to inherent noise, so the averages presented in this study are approximate.

We used the transparent scoring scheme proposed by Yvan et al. [58] to quantify the usability of each method, considering availability, documentation, code quality and publication in a peer-reviewed journal (Supplementary Data 4). Issues related to installation and code availability are widespread in the field of bioinformatics [59]. Although not directly tied to method performance, evaluating quality and user-friendliness is also crucial.

Score aggregation

To comprehensively assess the overall performance of each method, we aggregated the scores of the three different aspects mentioned above. For memory and time consumption, we first standardized the values of the different methods on the same dataset by transforming them to a standard normal distribution. We then used a probability density function to transform these values to the range of [0, 1]. Next, we calculated the arithmetic mean of the scores within each of the three aspect groups. Finally, to obtain a combined total score for the performance evaluation, we calculated the arithmetic mean of all the scores.

Method execution

Each execution of a method on a dataset was performed in a pipeline. All benchmarking tasks were conducted on an Ubuntu Linux server with dual CPUs. The server was equipped with Intel(R) Xeon(R) Gold 6148F CPUs running at 2.40 GHz, 192 GB of RAM, and Linux version 4.15.0-197-generic. Furthermore, the server was equipped with two Tesla P100 GPUs, each possessing 16 GB of memory.

Web server development

To facilitate researchers' access to the results of our comprehensive evaluation of various methods for selecting the most suitable approach for their research and to offer drug repositioning prediction services to the broader community, we have developed an online web tool called HN-DREP, which is freely accessible at http://hn-drep.lyhbio.com/home.

HN-DREP is a B/S (Browser/Server) architecture web application that follows a front-end/back-end separation model to enhance system usability, security and maintainability. The front-end of HN-DREP utilizes the React framework, a popular JavaScript library for developing user-friendly interfaces. Meanwhile, the back end relies on the industry-standard LNMP (Linux, Nginx, MySQL, PHP) technology stack, enhanced by the integration of the Redis cache and the ElasticSearch search engine to improve system responsiveness. Our entire system is deployed within Kubernetes for efficient management and scalability.

Key Points

- For the first time, a comprehensive evaluation of drug repositioning methods based on heterogeneous networks has been conducted, providing direction for method selection and new method development.
- A new evaluation workflow has been designed and implemented to rigorously evaluate drug repositioning methods based on heterogeneous networks, providing a standardized framework for researchers to evaluate new methods.
- To improve accessibility and usability, a user-friendly web interface has been developed that allows researchers to easily access evaluation results, download evaluated methods and datasets, and perform drug repositioning predictions for drugs or diseases of interest.

ACKNOWLEDGEMENTS

We extend our heartfelt thanks to Hongrui Duo for his invaluable contributions to the discussion and to Jin Zhang and Zunian Wang for their technical support.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 62101087), the China Postdoctoral Science Foundation (Grant No. 2021MD703942), the Chongqing Postdoctoral Research Project Special Funding (Grant No. 2021XM2016), the Science Foundation of Chongqing Municipal Commission of Education (Grant No. KJQN202100642 and KJQN202100538) and the innovation and entrepreneurship training program for college students in Chongqing (S202010617010).

AUTHORS' CONTRIBUTIONS

K.S., Y.L. and B.L. conceived and initiated this research, with Y.L. and Y.Y. writing the manuscript and evaluating all methods. Z.T., Y.W. and Q.M. conducted data collection and analysis, while M.B. and Y.Y. built the website. G.L. and M.L. generated the figures and tables and provided manuscript refinement. K.S. and B.L. supervised the entire project. All authors participated in discussions and finalized the manuscript.

DATA AVAILABILITY

The datasets used in this study and their related information are available on Zenodo at https://zenodo.org/record/8357512. To facilitate research and provide a benchmark for future evaluations of new methods and datasets, we have made available a Snakemake workflow named HN-DRES on GitHub (https://github.com/lyhbio/HN-DRES). Additionally, we have provided documentation to assist researchers in executing and extending this workflow.

To safeguard against the potential pitfalls of network or server downtime rendering HN-DREP unusable, we've taken the step of hosting the application's code on GitHub (https://github.com/lyhbio/HN-DREP/tree/master). For added convenience, we've also crafted a Docker image, streamlining the process for users to independently deploy HN-DREP (https://zenodo.org/records/10674628).

REFERENCES

1. Chong CR, Sullivan DJ. New uses for old drugs. *Nature* 2007;**448**:645–6.
2. Shim JS, Liu JO. Recent advances in drug repositioning for the discovery of new anticancer drugs. *Int J Biol Sci* 2014;**10**:654–63.
3. Pushpakom S, Iorio F, Eyers PA, *et al.* Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 2019;**18**:41–58.
4. Lotfi Shahreza M, Ghadiri N, Mousavi SR, *et al.* A review of network-based approaches to drug repositioning. *Brief Bioinform* 2018;**19**:878–92.
5. Yang M, Luo H, Li Y, Wang J. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* 2019;**35**:i455–63.
6. Cai L, Lu C, Xu J, *et al.* Drug repositioning based on the heterogeneous information fusion graph convolutional network. *Brief Bioinform* 2021;**22**:bbab319.
7. Zou J, Zheng M-W, Li G, Su ZG. Advanced systems biology methods in drug discovery and translational biomedicine. *Biomed Res Int* 2013;**2013**:742835 1, 8.
8. Lavecchia A, Cerchia C. In silico methods to address polypharmacology: current status, applications and future perspectives. *Drug Discov Today* 2016;**21**:288–98.

9. Xue H, Li J, Xie H, Wang Y. Review of drug repositioning approaches and resources. *Int J Biol Sci* 2018;**14**:1232–44.
10. Kim Y, Jung Y-S, Park J-H, *et al.* Drug-disease association prediction using heterogeneous networks for computational drug repositioning. *Biomolecules* 2022;**12**:1497.
11. Yang M, Wu G, Zhao Q, *et al.* Computational drug repositioning based on multi-similarities bilinear matrix factorization. *Brief Bioinform* 2021;**22**:bbaa267.
12. Zhao B-W, Hu L, You Z-H, *et al.* HINGRL: predicting drug–disease associations with graph representation learning on heterogeneous information networks. *Brief Bioinform* 2022;**23**:bbab515.
13. Luo H, Li M, Wang S, *et al.* Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics* 2018;**34**:1904–12.
14. Xie G, Xu H, Li J, *et al.* DRPADC: a novel drug repositioning algorithm predicting adaptive drugs for COVID-19. *Comput Chem Eng* 2022;**166**:107947.
15. Yang M, Huang L, Xu Y, *et al.* Heterogeneous graph inference with matrix completion for computational drug repositioning. *Bioinformatics* 2021;**36**:5456–64.
16. Yang X, Zamit L, Liu Y, He J. Additional Neural Matrix Factorization model for computational drug repositioning. *BMC Bioinformatics* 2019;**20**:423.
17. Rivas-Barragan D, Mubeen S, Guim Bernat F, *et al.* Drug2ways: reasoning over causal paths in biological networks for drug discovery. *PLoS Comput Biol* 2020;**16**:e1008464.
18. Li Z-C, Huang M-H, Zhong W-Q, *et al.* Identification of drug-target interaction from interactome network with 'guilt-by-association' principle and topology features. *Bioinformatics* 2016;**32**:1057–64.
19. Chiang AP, Butte AJ. Systematic evaluation of drug-disease relationships to identify leads for novel drug uses. *Clin Pharmacol Ther* 2009;**86**:507–10.
20. Yan Y, Yang M, Zhao H, *et al.* Drug repositioning based on multi-view learning with matrix completion. *Brief Bioinform* 2022;**23**:bbac054.
21. Zhao B-W, Su X-R, Hu P-W, *et al.* A geometric deep learning framework for drug repositioning over heterogeneous information networks. *Brief Bioinform* 2022;**23**:bbac384.
22. Wang X, Yan R. DDAPRED: a computational method for predicting drug repositioning using regularized logistic matrix factorization. *J Mol Model* 2020;**26**:60.
23. Gao C-Q, Zhou Y-K, Xin X-H, *et al.* DDA-SKF: predicting drug–disease associations using similarity kernel fusion. *Front Pharmacol* 2022;**12**:784171.
24. Zeng X, Zhu S, Liu X, *et al.* deepDR: a network-based deep learning approach to *in silico* drug repositioning. *Bioinformatics* 2019;**35**:5191–8.
25. Meng Y, Wang Y, Xu J, *et al.* Drug repositioning based on weighted local information augmented graph neural network. *Brief Bioinform* 2023;**25**:bbad431.
26. Zhang W, Xu H, Li X, *et al.* DRIMC: an improved drug repositioning approach using Bayesian inductive matrix completion. *Bioinformatics* 2020;**36**:2839–47.
27. Meng Y, Lu C, Jin M, *et al.* A weighted bilinear neural collaborative filtering approach for drug repositioning. *Brief Bioinform* 2022;**23**:bbab581.
28. Mongia A, Chouzenoux E, Majumdar A. Computational prediction of drug-disease association based on graph-regularized one bit matrix completion. *IEEE/ACM Trans Comput Biol Bioinform* 2022;**19**:3332–9.
29. Wang Y, Deng G, Zeng N, *et al.* Drug-disease association prediction based on Neighborhood information aggregation in neural networks. *IEEE Access* 2019;**7**:50581–7.
30. Chen H, Cheng F, Li J. iDrug: integration of drug repositioning and drug-target prediction via cross-network embedding. *PLoS Comput Biol* 2020;**16**:e1008040.
31. Yang M, Yang B, Duan G, Wang J. ITRPCA: a new model for computational drug repositioning based on improved tensor robust principal component analysis. *Front Genet* 2023;**14**:1271311.
32. Yu Z, Huang F, Zhao X, *et al.* Predicting drug–disease associations through layer attention graph convolutional network. *Brief Bioinform* 2021;**22**:bbaa243.
33. Luo H, Wang J, Li M, *et al.* Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* 2016;**32**:2664–71.
34. Sadeghi S, Lu J, Ngom A. A network-based drug repurposing method via non-negative matrix factorization. *Bioinformatics* 2022;**38**:1369–77.
35. Yang M, Luo H, Li Y, *et al.* Overlap matrix completion for predicting drug-associated indications. *PLoS Comput Biol* 2019;**15**:e1007541.
36. Zhang W, Yue X, Lin W, *et al.* Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinformatics* 2018;**19**:233.
37. Meng Y, Jin M, Tang X, Xu J. Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study. *Appl Soft Comput* 2021;**103**:107135.
38. Wang Y, Xiang J, Liu C, *et al.* Drug repositioning for SARS-CoV-2 by Gaussian kernel similarity bilinear matrix factorization. *Front Microbiol* 2022;**13**:1062281.
39. Xu J, Meng Y, Peng L, *et al.* Computational drug repositioning using similarity constrained weight regularization matrix factorization: a case of COVID-19. *J Cellular Molecular Medi* 2022;**26**:3772–82.
40. Wishart DS, Feunang YD, Guo AC, *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
41. Hamosh A, Scott AF, Amberger JS, *et al.* Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2004;**33**:D514–7.
42. Davis AP, Grondin CJ, Johnson RJ, *et al.* The comparative toxicogenomics database: update 2017. *Nucleic Acids Res* 2017;**45**:D972–8.
43. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol* 2011;**7**:496.
44. Martínez V, Navarro C, Cano C, *et al.* DrugNet: network-based drug-disease prioritization by integrating heterogeneous data. *Artif Intell Med* 2015;**63**:41–9.
45. Liang X, Zhang P, Yan L, *et al.* LRSSL: predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics* 2017;**33**:1187–96.
46. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics* 2014;**30**:2923–30.
47. Norel R, Rice JJ, Stolovitzky G. The self-assessment trap: can we all be better than average? *Mol Syst Biol* 2011;**7**:537.
48. Ye H, Liu Q, Wei J. Construction of drug network based on side effects and its application for drug repositioning. *PloS One* 2014;**9**:e87864.
49. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res* 2016;**44**:D1075–9.
50. Guha R. Chemical informatics functionality in *R*. *J Stat Soft* 2007;**18**:1–16.
51. Kastrin A, Ferk P, Leskošek B. Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning. *PloS One* 2018;**13**:e0196865.

52. Chung NC, Miasojedow B, Startek M, *et al*. Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. *BMC Bioinf* 2019;**20**:644.
53. Van Driel MA, Bruggeman J, Vriend G, *et al*. A text-mining analysis of the human phenome. *Eur J Hum Genet* 2006;**14**:535–42.
54. Schriml LM, Mitraka E, Munro J, *et al*. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res* 2019;**47**:D955–62.
55. Yu G, Wang L-G, Yan G-R, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 2015;**31**:608–9.
56. Weber LM, Saelens W, Cannoodt R, *et al*. Essential guidelines for computational method benchmarking. *Genome Biol* 2019;**20**:125.
57. Li W, Guo Q. Plotting receiver operating characteristic and precision–recall curves from presence and background data. *Ecol Evol* 2021;**11**:10192–206.
58. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol* 2019;**37**:547–54.
59. Mangul S, Mosqueiro T, Abdill RJ, *et al*. Challenges and recommendations to improve the installability and archival stability of omics computational tools. *PLoS Biol* 2019;**17**:e3000333.