

METHODOLOGY ARTICLE

Open Access



# Drug–target interaction prediction using unifying of graph regularized nuclear norm with bilinear factorization

Ali Ghanbari Sorkhi<sup>1\*</sup>, Zahra Abbasi<sup>2</sup>, Majid Iranpour Mobarakeh<sup>3</sup> and Jamshid Pirgazi<sup>1\*</sup>

\*Correspondence:

ali.ghanbari@mazust.ac.ir;

j.pirgazi@mazust.ac.ir

<sup>1</sup> Faculty of Electrical and Computer Engineering, University of Science

and Technology of Mazandaran, P.O. Box, 48518-78195 Behshahr, Iran

Full list of author information is available at the end of the article

## Abstract

**Background:** Wet-lab experiments for identification of interactions between drugs and target proteins are time-consuming, costly and labor-intensive. The use of computational prediction of drug–target interactions (DTIs), which is one of the significant points in drug discovery, has been considered by many researchers in recent years. It also reduces the search space of interactions by proposing potential interaction candidates.

**Results:** In this paper, a new approach based on unifying matrix factorization and nuclear norm minimization is proposed to find a low-rank interaction. In this combined method, to solve the low-rank matrix approximation, the terms in the DTI problem are used in such a way that the nuclear norm regularized problem is optimized by a bilinear factorization based on Rank-Restricted Soft Singular Value Decomposition (RRSSVD). In the proposed method, adjacencies between drugs and targets are encoded by graphs. Drug–target interaction, drug–drug similarity, target–target, and combination of similarities have also been used as input.

**Conclusions:** The proposed method is evaluated on four benchmark datasets known as Enzymes (E), Ion channels (ICs), G protein-coupled receptors (GPCRs) and nuclear receptors (NRs) based on AUC, AUPR, and time measure. The results show an improvement in the performance of the proposed method compared to the state-of-the-art techniques.

**Keywords:** Drug–target interaction, Computational prediction, Low-rank interaction, Drug discovery

## Background

The study of DTIs has been attracted many researchers' attention in the field of pharmaceutical science in recent years [1–4]. In this regard, many efforts have been made to investigate drug repositioning as well as the discovery of the interaction between new targets and existing drugs. DTI means binding a drug to a target location, that leads to a change in its behavior or function. On the other hand, the identification of DTIs minimizes the adverse side effects of drugs [1].



© The Author(s), 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Performing wet-lab experiments is a significant challenge in terms of cost, time and effort [5]. In this regard, Computational Prediction (CP) methods have been used in recent years [6]. In addition, there is ample evidence of Disease-Associated Microbes (DAM) as well as Long non-coding RNA (lncRNA)-Disease Associations (LDA) [7, 8]. Using traditional approaches of experiments to confirm these connections often requires a great deal of materials and time which are expected computational methods to be used to predict these associations. Many of these algorithms use profile-based methods (for example, NCPLP [7] in ADM and BLM-NPAI [8] in LDA) to predict these associations.

Despite the synthesis of many compounds, their target profiles and drug effects are still unidentified. Besides, there is no cure for many diseases and many new diseases are introduced each year. Therefore, much information has been gathered about various compound properties, features, responses and target proteins by researchers. The emergence of a large dataset has led to the use of CP with problems such as high dimensional, complex data, which indicates the need for efficient and robust algorithms in DTIs.

Computational methods for DTIs have been used by state-of-the-art researchers. Generally, three main categories can be introduced for computational methods in this application [1]. In the first category, the concept of that similar molecules tend to share similar properties and usually bind similar proteins, is used which these methods are called ligand-based approaches [6, 9, 10]. This approach predicts interactions using similarities between identical protein ligands. Since these ligand-based methods do not use sequence information of the proteins for prediction, it is possible that a novel interaction restricts to link between known ligands and protein families. On the other hand, the performance of these methods is highly dependent on known ligands, and if these ligands are low for a candidate protein, the performance of these methods is drastically reduced [1, 11].

In the second category, the 3D structures of drugs and proteins are used by a simulation to determine the interaction, known as docking approaches. The main problem is that the 3D structure of some proteins is not known [12, 13]. The third category is chemogenomic-based methods that uses information about drugs and targets simultaneously. This method has been considered by many researchers in recent years; furthermore, it can be used in a broad biological data, which are also used for the prediction of data from process information such as chemical structure graphs and genomic sequences for the drugs and targets from both sides of the drug and target simultaneously [14]. For this purpose, biological information that is available in public datasets can be used. This general method can be divided into two categories: Feature-based and similarity-based methods [15]. In the feature-based method, a supervised machine learning technique is used. In fact, in this method, feature vectors use sets of drug–target pairs with class labels that indicate the presence of interaction (positive instance) or no interaction (negative example). Also, it should be noted that negative samples are samples without non-interactions or unknown drug–target interactions [16–18]. In similarity-based methods, two matrices of similarity related to the drug and similarity of the target along with the interaction matrix are used which represent the interaction between the drug and the target [19–21]. These similarities are usually created for the drug by chemical structures and for the target by protein sequence alignment. Similarity-based methods have several apparent advantages [22]:

1. Feature-based approaches require a feature extraction or selection process which is complex and challenging, while similarity-based techniques do not require this process.
2. Computing similarity measures have already been expanded and used extensively such as chemical structure similarity for drugs and genomic sequence similarity for targets.
3. Similarity-based approaches can provide better performance in prediction since directly related to kernel methods.
4. Similarity matrices represent chemical space and genomic space derived from the relationships between drugs and genes, respectively.

These advantages demonstrate the superiority of similarity-based approaches over other approaches. Adjacency matrices are commonly used to represent drug-drug and target-target similarity.

Another point of view that can be introduced to categorize the available methods in DTI is methods which includes classification, network inference and matrix factorization groups. Classification-based models are divided into Local Classification Model (LCM) [23, 24] and Global Classification Model (GCM) [25, 26]. It is difficult to diagnose drugs (resp. targets) that interact with the same target (resp. drug), so the LCM is not able to show a link between targets or drugs [27]. Also, GCM cannot show the relationship between targets or drugs due to the complexity of similarity calculations based on the tensor product or high dimensional concatenate feature vectors. Overall, these models do not easily capture the underlying structure among drug–target pairs [28].

In recent years, many types of researches have been done based on deep learning in DTI [29–31]. A comprehensive deep learning library called DeepPurpose has been introduced for DTI prediction [29]. This library includes the implementation of 15 compound and protein encoders and more than 50 neural architectures with other beneficial features in DTI.

Convolution Neural Networks (CNNs) were used to obtain 1-dimensional representations of protein sequences (amino acid sequences) and simplified molecular input line-entry system (SMILES) compounds in [30]. The extraction features were claimed to show an appropriate representation of local dependencies or patterns and serve as a suitable input for a fully connected neural network (FCNN) for the binary classifier. The results show that the use of CNNs to obtain data display, as an alternative to traditional descriptors, improves performance in DTI.

A deep learning model based on DeepLSTM was developed to predict DTI in [31]. Position-Specific Scoring Matrices (PSSM) and Legendre Moment (LM) were used to extract the evolutionary features of proteins. The Sparse Principal Component Analysis (SPCA) was then used to compress the features of drugs and proteins in a uniform vector space.

It should also be noted that the use of deep learning also faces with similar problems; on the other hand, the use of deep learning has a large dataset for training, which unfortunately in these applications, providing data is expensive and time-consuming.

Interactions between drugs and targets show a significant relationship that is represented by a bi-partite network [32]. The information in this network is taken from

drug–target interactions. A bi-partite network is based on network inference (e.g., NBI [19]) which has transformed DTI prediction to link prediction between graph nodes. Two-step resource allocation is used by NBI to infer the potential links between nodes. Although, it just depends on the local or the first-order topology of nodes, it tends to completely bias the high-degree nodes [26]. In addition, NBI cannot predict the interaction between the target–drug pair without known accessible pathways in the network [32]. The heterogeneous network is a promising model. This network is made by a DTI network and two other networks which is produced by drug similarities and target similarities, respectively [28, 33].

The models based on matrix factorization, such as BMF2K [34], CMF [35], NRLMF [36] are good models to obtain structural information between drug–target interactions. Accordingly, drugs and targets are planned to a common low-rank feature space according to the drug similarity matrix and the target similarity matrix [28]. Two networks (dual-network  $L_{2,1}$ -collaborative matrix factorization) have been proposed to predict Drug–Disease Interactions (DDI), called the  $L_{2,1}$ -network matrix factor [37]. In this method, to achieve better results, the Gaussian interaction profile kernels and  $L_{2,1}$ -norm are presented. Moreover, the network similarities of drugs and diseases are combined with their chemical and semantic similarities. In order to identify potential links in biomedical bi-partite networks, a method called graph regularized generalized matrix factorization (GRGMF) is proposed to predict links [38]. For this purpose, a matrix factorization model is formulated to use latent patterns behind observed links. It is claimed that the results showed an improvement in the proposed method.

In this regard, factorization approaches can be used to predict DTIs [21]. In general, based on the reported experiences [6], matrix factorization methods have achieved the best results in DTI. Since there are few factors for DTI and these latent factors characterize drugs and targets, the DTI matrix can be converted into a latent factor matrix of drug and target. The DTI matrix is of low-rank which can be solved using matrix factorization. Matrix factorization is a bi-linear non-convex problem that there are no convergence guarantees [11]. Nuclear Norm Minimization (NNM) based methods have also been proposed to improve it. By shrinking all Singular Values (SVs) uniformly, the NNM is usually used to estimate the matrix rank. Despite the precise physical meanings of SVs, NNM cannot accurately estimate the matrix rank.

In this paper, unifying matrix factorization and NNM approaches combined with graph regularization penalties are proposed, which will be described in detail in the following sections. It should be noted that similarly to Mongia et al. study [11], a similarity matrix is used in this proposed method.

The strengths of the proposed method are as follows:

- Unifying nuclear norm with bilinear factorization is presented based on the similarity of drug–drug and target–target, which has caused the advantages of both methods to be combined.
- Rank-Restricted Soft Singular Value Decomposition method is used to optimize the nuclear norm minimization in the DTI problem.

- The performance of the proposed method based on AUC and AUPR measures had the best performance and also the results were suitable in datasets with different features.
- The time complexity of the running time is  $O(r^3 + n^b \log(n))$  for each iteration in the proposed algorithm. This complexity is polynomial, which has performed better than other new methods.
- The proposed method can be widely used in other applications of CP. Using the proposed method does not require a complex process.

The rest of the paper is organized as follows: In “**Methods**” section, the proposed method that includes a novel algorithm for DTI is introduced. The experimental setup and results, that include the introduction of datasets, evaluation criteria and comparison of methods, are presented in “**Discussion**” section. Finally, the conclusion and discussion are presented in “**Conclusion**” section.

## Methods

The interaction between the target and the drug is shown by the adjacency matrix  $X$ , where drugs are presented in rows and targets are presented in columns. The value of the matrix represents the interactions. Since all DTIs are not known, this matrix is partially observed. This is expressed as follow:

$$Y = P(X) \quad (1)$$

In Eq. (1),  $P$  is a sub-sampling operator which in this binary sampling matrix the value of 1 means that there is a known interaction and 0 means otherwise.  $Y$  is an available partially sampled DTI matrix. The purpose of this equation is to estimate the matrix  $X$  given  $Y$  and known  $P$ .  $X$  is a low-rank that needs to be recovered. Equation (2) can be used for this purpose.

$$\min_X \text{rank}(X) \text{ such that } Y = P(X) \quad (2)$$

Low-Rank Matrix Approximation (LRMA) has been used in many practical cases that have low rank properties, so in recent years it has attracted considerable interest in different areas, such as computer vision and machine learning [39–42]. In general, LRMA methods are divided into two categories, the low rank matrix factorization (LRMF) [43–46] and the rank minimization methods [47–49]. The purpose of the LRMF concerning the input matrix  $Y$  is to factorize it to the product of two low rank matrices that can be used to reconstruct the low rank matrix  $X$  with exceptional fidelity. A variety of LRMF-based methods, such as classical Singular Value Decomposition (SVD) under ‘ $\mathcal{L}_2$  – norm’ [50, 51], robust LRMF methods under ‘ $\mathcal{L}_1$  – norm’ [52, 53] and other probabilistic methods have been proposed [54, 55]. The problem of Low rank models for recovering a rank- $k$  matrix  $Z$  can be expressed by minimizing Eq. (3).

$$\min_Z f(Y - Z) \text{ subject to } \text{rank}(Z) = k \quad (3)$$

where  $f(\cdot)$  defines a loss function. The rank limitation in Eq. (3) has typically been imposed by a factorization  $Z = AB^T$ , as

$$\min_{A, B} f(Y - AB^T) \quad (4)$$

based on its intractability. It has been proven that when the loss function is the Least Squares (LS) loss, i.e.,  $f(Y - AB^T) = \|Y - AB^T\|_F^2$ , then Eq. (4) does not have local minima and a closed form solution can be obtained via the SVD of  $Y$  [56]. One of the disadvantages of this factorization approach is highly susceptibility of the LS loss to outliers and the presence of missing data in  $Y$  results in local minima. Factorization with missing data is a NP-Hard problem [57], while outliers can be addressed with robust loss functions [58, 59].

In DTI, matrix  $X$  can be converted into two matrices as follows:

$$X_{M \times N} = A_{M \times k} B_{k \times N}, \quad k \ll (m, n) \quad (5)$$

Here  $M$  and  $N$  are the numbers of drugs and targets, respectively and  $k$  is the presumed rank of the matrix. The Eq. (4) for DTI is expressed as Eq. (6).

$$\min_{A, B} f(Y - P(AB^T)) \quad (6)$$

As mentioned earlier, the second category of LRMA methods is based on rank minimization. These methods by setting an additional rank constraint on the estimated matrix can reconstruct the data matrix. Direct rank minimization is challenging to solve because these are NP-hard. To solve this type of problem, the NNM methodology is used. In this methodology, the problem is generally solved by replacement minimizing the nuclear norm of the estimated matrix that is a convex relaxation of minimizing the matrix rank.  $\|X\|_*$  is the nuclear norm of matrix  $X$ . For example,  $\|X\|_* = \sum_i \sigma_i$  is the nuclear norm which is the sum of its SV that  $\sigma_i$  represents the  $i$ -th SV of the matrix  $X$ . NNM attempts to recover matrix  $X$ , actual low rank, by minimizing  $\|X\|_*$  from degraded observation matrix  $Y$ . In recent years, NNM-based methods have been used in many applications such as video denoising [60], background extraction [61], data recovery [62] and subspace clustering [63, 64]. The matrix rank can be recovered under the conditions of the limited and theoretic warranty. However, in some applications, it acts the various rank components equally, and therefore it cannot be precise enough to estimate the matrix rank. Thus several methods have been proposed to improve NNM performance [11, 65].

For noisy input, by solving the NNM problem, the inherent low rank reconstruction can be achieved with a high probability. Also, the Nuclear Norm Proximal (NNP) is also represented by the following equation:

$$\min_Z f(Y - Z) + \lambda \|Z\|_* \quad (7)$$

By using a soft threshold process on the SV of the observation matrix, it can be easily solved in closed form:

$$\hat{X} = US_{\frac{\lambda}{2}}(\Sigma)V^T \quad (8)$$

In this equation,  $Y = U\Sigma V^T$  is the SVD of  $Y$ , the soft thresholding function on diagonal matrix  $\Sigma$  with parameter  $\frac{\lambda}{2}$  is indicated by  $S_{\frac{\lambda}{2}}(\Sigma)$ . For each diagonal element  $\Sigma_{ii}$  in  $\Sigma$ , there is:

$$S_{\frac{\lambda}{2}}(\Sigma)_{ii} = \max\left(\Sigma_{ii} - \frac{\lambda}{2}, 0\right) \quad (9)$$

$\lambda$  is considered as a trade-off parameter between the loss function and the low-rank regularization, that is created through the nuclear norm. These models have generalized the use of low-rank compared to many applications, where  $Z$  is low rank but has no a priori [40, 66].

In addition to having convexity and theoretic guidance of the  $\lambda$  [40], these models also have multiple drawbacks.

To show how to create a determined rank in  $Z$ , by setting  $\lambda$  [67],  $Z$  has a predetermined rank. It usually gives more undesirable results than its direct usage in Eq. (2). Additionally, the “kernel trick” cannot be used because access to the factorization of  $Z$  in Eq. (3) is not available. Also, Eq. (3) is a Semidefinite Program (SDP) and Off-the-shelf SDP optimizers are suitable for low-middle dimensional optimization (i.e., hundreds of variables) and are not amenable for large scale datasets with the high dimension.

To deal with the limitations, this paper uses a robust method called Rank-Restricted Soft SVD (RRSSVD) based on Hastie et al. study [68] for the DTIs. In the following, we will describe this method in this application. Based on [69], the nuclear norm can be expressed as follows:

$$\|X\|_* = \min_{A,B} \frac{1}{2} (\|A\|_F^2 + \|B\|_F^2) \text{ subject to } X = AB^T \quad (10)$$

In this section, the relationship between factorization and nuclear norm approaches is used based on the method presented in [67], that bridges the gap between two methods is presented in Eq. (11).

$$\min_{A,B} f(Y - AB^T) + \frac{\lambda}{2} (\|A\|_F^2 + \|B\|_F^2) \quad (11)$$

Equation (11) for the DTI problem can be expressed as follows:

$$\min_{A,B} \|Y - P(AB^T)\|_F^2 + \frac{\lambda}{2} (\|A\|_F^2 + \|B\|_F^2) \quad (12)$$

In this paper, Eq. (12) is used to solve Eq. (13).

$$\min_Z \|Y - P(Z)\|_F^2 + \lambda \|Z\|_* \quad (13)$$

To solve the Eq. (13), Algorithm 1 is used based on the RRSSVD method in DTI prediction. In this algorithm, theorems 1 and 2 [68] are used for DTI.

**Theorem 1** For the optimization problem (14), where  $Y_{m \times n}$  is a fully observed matrix and  $0 < r \leq \min(m, n)$ .



$$\min_{Z : \text{rank}(Z) \leq r} F_\lambda(Z) := \frac{1}{2} \|Y - Z\|_F^2 + \lambda \|Z\|_* \quad (14)$$

a solution provided by

$$\tilde{Z} = U_r S_\lambda(D_r) V_r^T \quad (15)$$

where the rank- $r$  SVD of  $Y$  is  $U_r D_r V_r^T$  and  $S_\lambda(D_r) = \text{diag}[(\sigma_1 - \lambda)_+, \dots, (\sigma_r - \lambda)_+]$ .

**Theorem 2** For the optimization problem (14), where  $Y_{m \times n}$  is a fully observed matrix and  $0 < r \leq \min(m, n)$ .

$$\min_{A_{m \times r}, B_{n \times r}} \frac{1}{2} \|Y - AB^T\|_F^2 + \frac{\lambda}{2} (\|A\|_F^2 + \|B\|_F^2) \quad (16)$$

a solution provided by  $\tilde{A} = U_r S_\lambda(D_r)^{\frac{1}{2}}$  and  $\tilde{B} = V_r S_\lambda(D_r)^{\frac{1}{2}}$ , and all solutions satisfy  $\tilde{A}\tilde{B}^T = \tilde{Z}$ , where  $\tilde{Z}$  is as given in (15).

Finally, it should be noted that relative change in the Frobenius norm has been used to check convergence. Equation (17) is used to calculate it. This equation is based on a pair of iterates  $(U, D^2, V)$  (old) and  $(\tilde{U}, \tilde{D}^2, \tilde{V})$  (new).

$$\nabla F = \frac{\|UD^2V^T - \tilde{U}\tilde{D}^2\tilde{V}^T\|_F^2}{\|UD^2V^T\|_F^2} = \frac{\text{tr}(D^4) + \text{tr}(\tilde{D}^4) - 2\text{tr}(D^2U^T\tilde{U}\tilde{D}^2\tilde{V}^TV)}{\text{tr}(D^4)} \quad (17)$$

By this algorithm,  $\tilde{Z} = \tilde{A}\tilde{B}^T$  is considered as the output.

---

**Algorithm 1.** The proposed RRSSVD algorithm for DTI

---

Procedure RRSSVD\_DTI( $Y, P, \lambda$ )

Initialize  $A = UD$  where  $U_{m \times r}$  is a randomly chosen matrix with orthonormal columns and  $D = I_r$ , the  $r \times r$  identity matrix.

For Loop (until convergence of  $AB^T$ )

a. Given  $A$ , solve for  $B$ :

$$\min_B \|Y - P(AB^T)\|_F^2 + \lambda \|B\|_F^2$$

This is a multiresponse ridge regression, with a solution

$$\tilde{B}^T = (D^2 + \lambda I)^{-1} D U^T P(Y)$$

This is simply matrix multiplication followed by coordinate-wise shrinkage.

b. Compute the SVD of  $\tilde{B}D = \tilde{V}\tilde{D}^2R^T$ , and let  $V \leftarrow \tilde{V}$ ,  $D \leftarrow \tilde{D}$ , and  $B = VD$

c. Given  $B$ , solve for  $A$ :

$$\min_A \|Y - P(AB^T)\|_F^2 + \lambda \|A\|_F^2$$

This is also a multiresponse ridge regression, with solution

$$\tilde{A} = P(Y)VD(D^2 + \lambda I)^{-1}$$

Again matrix multiplication followed by coordinate-wise shrinkage

d. Compute the SVD of  $\tilde{A}D = \tilde{U}\tilde{D}^2R^T$ , and let  $U \leftarrow \tilde{U}$ ,  $D \leftarrow \tilde{D}$ , and  $A = UD$

End Loop

Compute  $M = P(Y)V$ , and then it's SVD:  $M = UD_\sigma R^T$ . Then output  $U, V \leftarrow VR$ , and  $S_\lambda(D_\sigma) = \text{diag}[(\sigma_1 - \lambda)_+, \dots, (\sigma_r - \lambda)_+]$

---

As mentioned in this paper, the adjacency matrix is used that represents the interaction matrix between targets and drugs. In this matrix, if there is a known interaction between the drug (dt) and the target ( $t_j$ ), the value is 1 and otherwise the value is zero. In this article, in addition to the interaction matrix, a drug similarity matrix ( $S_d$ ) and a target similarity matrix ( $S_t$ ) are used. With the number of substructures shared in the chemical structure between the two drugs, the SIMCOMP



introduced in [70] is used. In fact,  $S_d$  indicates the similarity of the chemical structure of the drug pair. Similarly,  $S_t$  represents the degree of similarity between the two proteins, which is calculated from the similarity of the genome sequence based on the amino acid sequence of the target protein. It should be noted that normalized Smith-Waterman [71] has been used to calculate this.

In addition to the use of the similarity matrix introduced, there are four other types of similarity matrices such as Cosine ( $S^{cos}$ ), Correlation ( $S^{cor}$ ), Hamming ( $S^{ham}$ ) and Jaccard ( $S^{jac}$ ) which are used to predict DTI [11]. This paper also uses five similarity matrices calculated using the drug–target interaction matrix. The similarity matrices are used for DTI prediction, which by this method; Eq. (13) is expressed as follows:

$$\min_Z \|Y - P(Z)\|_F^2 + \lambda \|Z\|_* + \alpha_1 \text{Tr}\left(Z^T \sum_{i=1}^{nsim} L_d^i X\right) + \alpha_2 \text{Tr}\left(Z^T \sum_{i=1}^{nsim} L_t^i X^T\right) \quad (18)$$

In Eq. (18),  $\alpha_1 > 0$  and  $\alpha_2 > 0$  are the balancing parameters,  $\text{Tr}(\cdot)$  is the trace of a matrix,  $nsim$  indicates the number of similarity matrices (similar reference [11]  $nsim = 5$ ).  $L_d$  and  $L_t$  are the graph Laplacians [72] for  $S_d$  and  $S_t$ , where  $L_d = D_d - S_d$  and  $L_t = D_t - S_t$  are computed, respectively.  $D_d$  and  $D_t$  are degree matrices for drugs and targets that are calculated as  $D_d^{ii} = \sum_j S_d^{ij}$  and  $D_t^{ii} = \sum_j S_t^{ij}$ .

As shown in Algorithm 2, a method proposed by Mongia et al. [11] is used to solve this equation. It should be noted that this algorithm uses the RRSVD\_DTI method that is proposed for DTI prediction in this paper.

**Algorithm 2.** New MGRNNM (MGRNNM2) method based on reference [11] combined with proposed RRSVD\_DTI method

---

```

Procedure MGRNNM2(M,A,Sdcom,Stcom)
  Sparsify: Sdcom, Stcom
  Initialize: λ, α1, α2, v1, v2, Ldcom, Ltcom, Y = M, Z = MT
  AA ←  $\begin{pmatrix} A \\ \sqrt{v_1}I \\ \sqrt{v_2}I \end{pmatrix}$ 
  For loop, iterate (k)
    YYk ←  $\begin{pmatrix} M \\ \sqrt{v_1}Z^T \\ \sqrt{v_2}Y \end{pmatrix}$ 
    Xk ← RRSVD_DTI(YYk, AA, λ)
    Yk ← solve - sylvester(v1l, α1Ldcom, v1Xk)
    Zk ← solve - sylvester(v2l, α2Ltcom, v2Xk)
  End Loop
  
```

---

In Algorithm 2,  $S_d^{com} = S_d + S_d^{cos} + S_d^{cor} + S_d^{ham} + S_d^{jac} = \sum_{i=1}^{nsim} S_d^i$  and  $S_t^{com} = S_t + S_t^{cos} + S_t^{cor} + S_t^{ham} + S_t^{jac} = \sum_{i=1}^{nsim} S_t^i$  represent the combined similarity for drug and target,  $D_d^{com} = \text{diag}\left(\sum_j S_d^{com}\right)$  and  $D_t^{com} = \text{diag}\left(\sum_j S_t^{com}\right)$  represent the combined degree matrix for the drug and target also  $L_d^{com} = D_d^{com} - S_d^{com}$  and  $L_t^{com} = D_t^{com} - S_t^{com}$  represent the combined Laplacian matrix for the drug and target, respectively.

## Result

In this section, the experiments and results based on the proposed method are analyzed separately. In the first step, datasets and how to divide them into training and testing sets are presented. All experiments and extracted parameters are performed separately for each dataset under different validation settings. The following compares the proposed method with other new methods based on AUC, AUPR and time criteria.

## Dataset description

Reference [14] examines information on drug and target proteins interactions for the public databases; KEGG BRITE [73], BRENDA [74] SuperTarget [75] and DrugBank [76]. In this paper, similar to [2, 11, 14], four benchmark datasets are used, which are from four different classes of target proteins. In fact, these benchmarks are simulated from public databases. The following is a description of these datasets:

- **Enzymes (Es):** In this dataset, 445 drugs, 664 targets and 2926 interactions have been extracted.
- **Ion channels (ICs):** In this dataset, 201 drugs, 204 targets and 1476 interactions have been extracted
- **G protein-coupled receptors (GPCRs):** In this dataset, 223 drugs, 95 targets and 635 interactions were extracted
- **Nuclear receptors (NRs):** In this dataset, 54 drugs, 26 targets and 90 interactions have been extracted

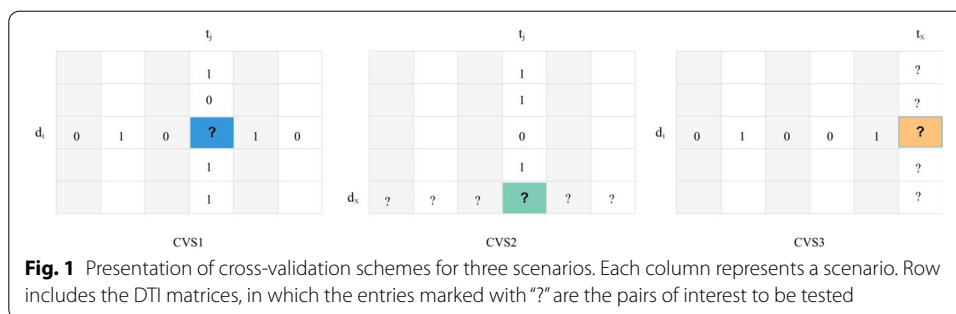
It should be noted that these datasets are simulated from public databases which at the link: <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/> are publicly available.

## Experimental setup

In this section, the setting of datasets is based on recent work done on the DTI problem. Three cross-validation settings (CVS) as named CVS1, CVS2 and CVS3 are introduced [6]. In CVS1, standard setting for evaluation, the target-drug pairs for the test set were randomly selected for prediction. In CVS2 and CVS3, settings are performed to evaluate the ability of methods to predict interactions for novel drugs (i.e., drugs for which no interaction information is available) and novel targets, respectively. It can be pointed out that in CVS2, entire drug profiles and CVS3, total target profiles are selected as a test set.

When at least one DTI is known for  $d_i$  and  $t_j$  respectively in the training data the CVS1 predicts the unknown pair  $(d_i, t_j)$ . To prevent using the pairs, CV used the pairs between the drugs having at least 2 targets and the targets interacting with at least 2 drugs, which should be used in three other scenarios. Some of these pairs are selected randomly for testing in each round of CV and the union of the rest of them and other entries are used for training.

However, when there are no DTIs for observation of new drugs and new targets in the training data, CVS2 and CVS3 predict new drugs and new targets respectively.



Performance of CV on drugs in CSV2, where the rows corresponding to drugs are randomly blinded for testing and the remaining rows are used for training. Performance of CV on targets in CSV3, where the columns (accounting for targets) are randomly blinded for testing and the resting columns are used for training.

We have made various tasks of CV under 3 scenarios shown in Fig. 1, respectively [28].

Similarly to the method presented by Mongia et al. [11], tenfold cross-validation (CV) is used, where data was divided into tenfolds and out of those 10 folds, one was selected as a test set while the remaining ninefolds were considered as a training set. In experiments, 5 repetitions of tenfold CV for each of the methods under three CVS are performed. For a more accurate evaluation, in each repetition, all CVSs for each dataset are similar to Mongia and Ezzat study [6, 11].

### Evaluation metrics

To evaluate the performance of the proposed method, Area Under the ROC Curve (AUC) and Area Under the Precision-Recall curve (AUPR) criteria based on Mongia and Ezzat study [6, 11] have been used. In the following, the requirements are introduced:

- AUC is a famous quality measure of ranking performance. It uses the ROC curve, which is a graphical plot that illustrates the diagnostic ability with a positive rate for a method as a function of the false-positive rate. AUC measures all two-dimensional areas below the ROC curve. It is also used as a measure of classification performance, aggregating over decision thresholds. Interpreting AUC shows a better model that is a random positive example more highly than a random negative example.
- AUPR is another measure which is used to evaluate the performance of DTI methods in this paper. It uses the precision–recall curve, which is a ratio of true positives plot that illustrates the positive predictions for each given recall rate. AUPR performance evaluation shows this area under the Precision-Recall curve punishes more false positives than AUC. The AUPR offers a quantitative assessment of the separation of true interactions from true non-interactions among predicted scores. For this reason, due to few true drug–target interactions, AUPR is a more important qualitative scale than AUC which finds true drug–target interactions among prediction scores.

**Table 1** Extracted parameters for the proposed method and [11]

Validation setting	Datasets	Parameters						
		$P$	$\lambda$	$\alpha_1$	$\alpha_2$	$v_1$	$v_2$	$r$
CVS1	NR	2	0.1	0.1	0.01	0.1	0.1	4
	GPCR	2	0.5	0.5	0.1	0.1	0.5	4
	IC	5	0.1	0.01	0.1	0.1	0.01	4
	E	5	0.1	0.01	0.1	0.1	0.1	4
CVS2	NR	2	0.01	0.01	0.01	0.01	0.1	4
	GPCR	2	0.01	0.1	0.01	0.01	0.01	4
	IC	5	0.1	0.01	0.1	0.1	0.1	4
	E	5	0.01	0.1	0.01	0.1	0.01	4
CVS3	NR	2	0.01	0.01	0.01	0.01	0.1	4
	GPCR	2	0.1	0.01	0.01	0.1	0.1	4
	IC	2	0.1	0.01	0.1	0.01	0.1	4
	E	2	0.1	0.01	0.1	0.01	0.01	4

### Parameter settings

In this paper, cross-validation on the training set is used to set the parameters of the proposed method. In fact, experiments are designed under each cross-validation setting to find the best parameter for each dataset. The considered parameters are  $P, \lambda, \alpha_1, \alpha_2, v_1, v_2, r$ . It should be noted that the ranges for  $\lambda, \alpha_1, \alpha_2, v_1, v_2$  and  $P, r$  parameters are considered (0,1) and (1,10), respectively. All the extracted parameters are shown in Table 1.

It is necessary to mention, implementations<sup>1</sup> were performed in MATLAB programming language on hardware configuration, 4G memory, and Core i7 M620 2.6 GHz CPU.

### Interaction prediction

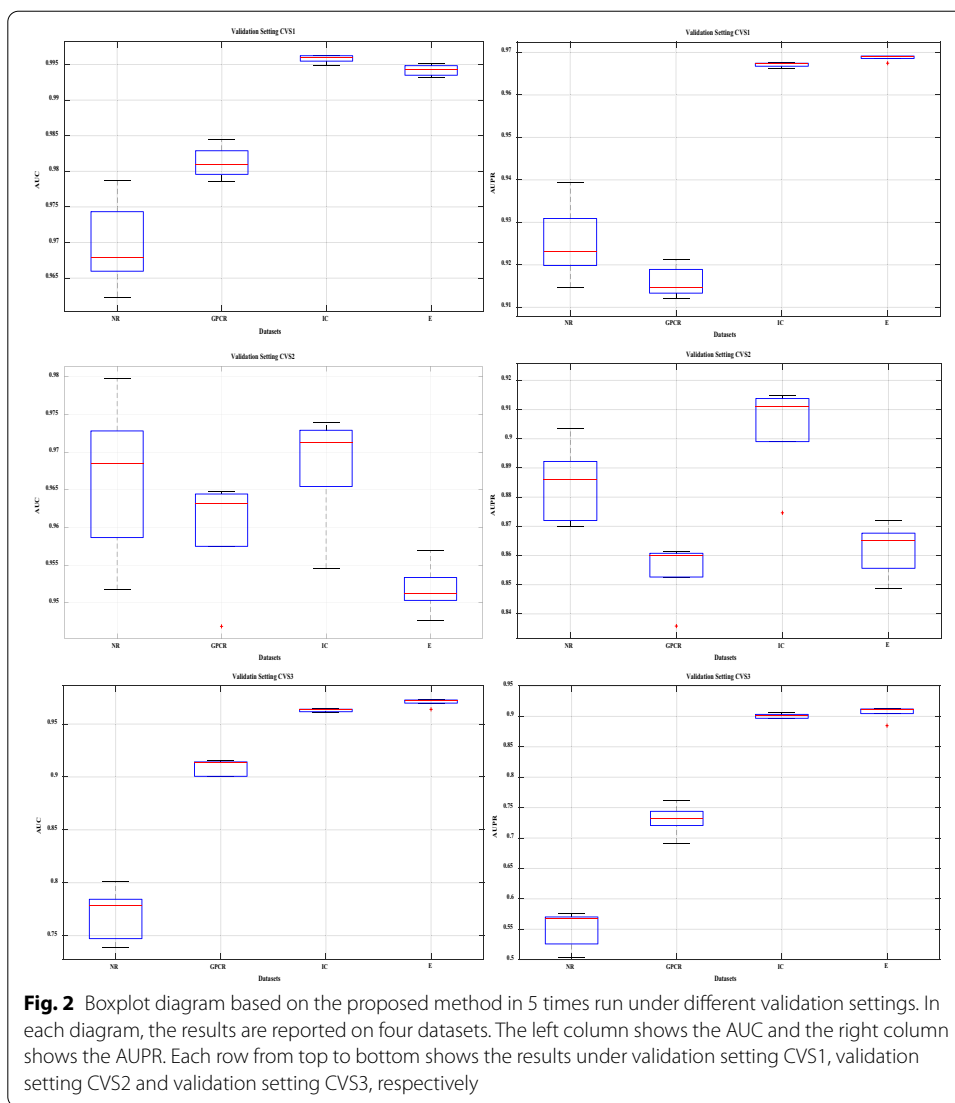
In Fig. 2, the results obtained in each validation set are shown by boxplot. In fact, this diagram is drawn based on five different runs on four databases. Two criteria, AUC and AUPR, have been considered in drawing this diagram. The results show the performance of the proposed method which is appropriate in each run and is without outliers.

### Comparison with the others methods

In this paper, to evaluate the proposed method, experiments have been designed to compare this method with 6 state-of-the-art methods introduced in recent years. In the following, we will describe these methods.

Weighted Graph Regularized Matrix Factorization (WGRMF) method [20] was introduced in 2016. In this method, since the data are located on or close to low-dimensional non-linear manifolds, two methods of matrix factorization are proposed, which in these methods, graph regularization is used. Also, a preprocessing step has been presented to improve the predictions of a “new drug” and a “new target” by introducing intermediate interaction likelihood scores.

<sup>1</sup> <https://github.com/ali289/DTI-GRNNwBF/>



In 2019, an improved graph regularized matrix factorization (GRMF) method was proposed to learn DTI flow patterns by combining the matrix analysis method called  $L_{1,2}$ -GRMF [77]. In this method, WKNKN for preprocessing is used to improve prediction accuracy.

The Collaborative Matrix Factorization (CMF) method [35] was introduced in 2013. The main idea of this method is to use more than one target and drug similarity matrix. In this method, a weighted matrix for the automatic selection of similarities is estimated to improve DTI prediction.

Subsequently, a factor model called Multiple Similarities Collaborative Matrix Factorization (MSCMF) is proposed in which drugs and targets are projected in a common low-rank feature space. Finally, these two low-rank matrices and weights associated with similarity matrices are estimated by an alternating least squares algorithm.

Regularized Least Square Weighted Nearest Neighbor profile (RLS-WNN) method [25] was introduced in 2013. In this method, a simple weighted nearest neighbor

procedure is introduced which is claimed that the procedure has performed well in DTI prediction, and to improve on previous work, this procedure is combined with the recent machine learning method.

Multi Graph Regularized Nuclear Norm Minimization (MGRNNM) method [11] was proposed in 2020. In this method, a new framework for predicting DTI from three inputs; known drug–target interaction network, similarities over drugs and those over targets is proposed. A method for finding a low-rank interaction matrix has been introduced. This matrix is made up of graphs that represent the proximities of drugs and targets. This paper proposes to capture the proximities exhaustively in predicting DTI, various multiple drug–drug similarities and target–target similarities as multiple graph Laplacian (over drugs / targets) regularization terms be used.

Four references are introduced specifically for the DTI prediction task. In the following, two references are introduced as baseline references.

Matrix completion (MC) method [78] was introduced in 2011. The paper focuses on solving matrix completion problems. In this regard, a non-convex optimization problem is proposed to solve the matrix completion. The proposed method is a variant of convex nuclear-norm minimization, with a fast numerical algorithm to solve it.

Matrix Completion on Graphs (MCG) method [79] was introduced in 2014. This paper introduces a novel matrix completion model for several real-world applications such as recommender systems. In this new model, the proximity information is used. It is stated that the purpose of this method is to find a low-rank solution created by the proximities of rows and columns. It should be noted that these proximities are encoded by graphs.

Neighborhood Regularized Logistic Matrix Factorization (NRLMF) method [36] was introduced in 2016. This method is based on the possibility of interacting a drug with a target through the logistic matrix factorization. NRLMF is more important in drug–target interaction pairs (positive observations) than in unknown pairs (negative observations). Because positive observations have already been experimentally confirmed, they are usually more reliable. For this reason, the local structure of drug–target interaction data has also been used through neighborhood adjustment to achieve better predictive accuracy.

DDR method was introduced in 2018 [80]. DDR works by using multiple similarities between drugs and considerable similarities between target proteins through a heterogeneous graph containing known DTIs.

Triple Matrix Factorization-based model (TMF) [28] was introduced in 2018. This model shows a new sight for the effective mechanism of DTIs by indicating prevailing features. TMF assesses the predictions on four benchmark datasets over different screening scenarios which represent its considerable priority.

In this section, a comparison between the proposed method and the current prediction methods is performed. AUC and AUPR criteria were used to evaluate the performance. Tables 2, 3, 4 and 6 show the results found on AUPR under validation setting; CVS1, CVS1 and CVS1 in four data sets, respectively. Tables 3, 5, 6 and 7 also show the results based on AUC under validation setting; CVS1, CVS1 and CVS1 in four datasets, respectively. In these tables, the best result are shown in bold. As shown in the tables, the proposed method performed well in all four data sets and all three

**Table 2** AUPR results for interaction prediction under validation setting CVS1

Method	Datasets			
	E	IC	GPCR	NR
MGRNNM [11]	0.9660 (0.0006)	0.9585 (0.0013)	0.8515 (0.0033)	0.8791 (0.0019)
SMGRNNM [11]	0.9014 (0.0018)	0.9298 (0.0026)	0.7483 (0.0039)	0.6408 (0.0234)
MC [78]	0.7882 (0.0022)	0.8868 (0.0028)	0.6481 (0.0116)	0.3950 (0.0298)
MCG [79]	0.7621 (0.0025)	0.8346 (0.0025)	0.5956 (0.0102)	0.4558 (0.0202)
WGRMF [20]	0.8768 (0.0020)	0.9225 (0.0022)	0.7370 (0.0024)	0.6016 (0.0378)
WKNKN + L <sub>1,2</sub> -GRMF [77]	–	–	–	–
RLS-WNN [25]	0.8093 (0.0045)	0.8459 (0.0106)	0.6933 (0.0226)	0.7072 (0.0290)
CMF [35]	0.8837 (0.0026)	0.9373 (0.0019)	0.7543 (0.0017)	0.6383 (0.0149)
NRLMF [36]	0.892 (0.006)	0.906 (0.008)	0.749 (0.015)	0.728 (0.041)
TMF [28]	0.952 (0.002)	0.952 (0.002)	0.844 (0.006)	0.811 (0.035)
DDR [80]	0.92	0.92	0.79	0.83
Proposed method	<b>0.9687 (0.0007)</b>	<b>0.9670 (0.0005)</b>	<b>0.9159 (0.0036)</b>	<b>0.9253 (0.0091)</b>

**Table 3** AUC results for interaction prediction under validation setting CVS1

Method	Datasets			
	E	IC	GPCR	NR
MGRNNM [11]	<b>0.9955 (0.0003)</b>	0.9947 (0.0004)	0.9785 (0.0020)	0.9660 (0.0056)
SMGRNNM [11]	0.9798 (0.0004)	0.9829 (0.0012)	0.9531 (0.0028)	0.9083 (0.0058)
MC [78]	0.9596 (0.0015)	0.9415 (0.0015)	0.8110 (0.0055)	0.5882 (0.0253)
MCG [79]	0.8753 (0.0023)	0.9539 (0.0010)	0.8977 (0.0047)	0.8315 (0.0165)
WGRMF [20]	0.9647 (0.0013)	0.9747 (0.0022)	0.9432 (0.0010)	0.8892 (0.0153)
WKNKN + L <sub>1,2</sub> -GRMF [77]	–	–	–	–
RLS-WNN [25]	0.9635 (0.0014)	0.9786 (0.0026)	0.9458 (0.0044)	0.9329 (0.0114)
CMF [35]	0.9705 (0.0013)	0.9832 (0.0008)	0.9493 (0.0031)	0.8679 (0.0124)
NRLMF [36]	0.987 (0.001)	0.989 (0.001)	0.969 (0.004)	0.950 (0.011)
TMF [28]	0.989 (0.001)	0.989 (0.001)	0.9830 (0.003)	<b>0.978 (0.008)</b>
DDR [80]	0.97	0.98	0.96	0.92
Proposed method	0.9941 (0.0008)	<b>0.9958 (0.0005)</b>	<b>0.9832 (0.0022)</b>	0.9698 (0.0062)

validation sets. It should be noted that the SMGRNNM method is the same method presented in the Mongia et al. study [11], which uses only the standard similarity matrices.

### Time complexity

In this section, the time complexity of the proposed algorithm is compared with MGRNNM method [11] as one of the best methods presented in DTI. Since Algorithm 2 is an iterative solution, in this paper, similar to MGRNNM, the time complexity is calculated for each iteration. In the proposed method and MGRNNM, in each iteration two.

Sylvester equations and one NNM are solved. According to Kirrinnis et al. study [81], the complexity of solving the Sylvester equation is equal to  $O(n^b \log(n))$ , in which the parameter  $b$  is between 2 and 3. To solve NNM in MGRNNM, the SVS method is used which the complexity of this algorithm is equal to  $O(n^3)$  in each iteration. In fact,



**Table 4** AUPR results for interaction prediction under validation setting CVS2

Method	Datasets			
	E	IC	GPCR	NR
MGRNNM [11]	0.8603 (0.0095)	0.9026 (0.0197)	0.8538 (0.0112)	0.8773 (0.0125)
SMGRNNM [11]	0.4089 (0.0104)	0.3650 (0.0178)	0.4175 (0.0076)	0.5620 (0.0262)
MC [78]	0.0114 (0.0005)	0.0473 (0.0035)	0.0404 (0.0017)	0.1120 (0.0206)
MCG [79]	0.0457 (0.0008)	0.0925 (0.0013)	0.1091 (0.0044)	0.2404 (0.0337)
WGRMF [20]	0.4019 (0.0128)	0.3666 (0.0169)	0.4247 (0.0113)	0.5695 (0.0136)
WKNKN + L <sub>1,2</sub> -GRMF [77]	0.386 (0.013)	0.356 (0.0012)	0.394 (0.007)	0.573 (0.011)
RLS-WNN [25]	0.2409 (0.0272)	0.3090 (0.0200)	0.3463 (0.0106)	0.5373 (0.0216)
CMF [35]	0.3848 (0.0094)	0.3538 (0.0137)	0.4059 (0.0104)	0.5203 (0.0250)
NRLMF [36]	0.358 (0.040)	0.344 (0.033)	0.364 (0.023)	0.545 (0.054)
TMF [28]	0.438 (0.016)	0.376 (0.017)	0.428 (0.011)	0.541 (0.033)
DDR [80]	0.73	0.69	0.63	0.71
Proposed method	<b>0.8619 (0.0089)</b>	<b>0.9042 (0.0167)</b>	<b>0.8552(0.109)</b>	<b>0.8841 (0.0135)</b>

**Table 5** AUC results for interaction prediction under validation setting CVS2

Method	Datasets			
	E	IC	GPCR	NR
MGRNNM [11]	0.9460 (0.0033)	<b>0.9714 (0.0095)</b>	0.9567 (0.0084)	0.9533 (0.0127)
SMGRNNM [11]	0.8260 (0.0108)	0.7913 (0.0090)	0.8805 (0.0024)	0.8452 (0.0215)
MC [78]	0.5060 (0.0090)	0.5512 (0.0034)	0.5855 (0.0039)	0.5294 (0.0200)
MCG [79]	0.7413 (0.0118)	0.7196 (0.0071)	0.7745 (0.0027)	0.6992 (0.0244)
WGRMF [20]	0.7982 (0.0144)	0.7902 (0.0149)	0.8800 (0.0025)	0.8615 (0.0244)
WKNKN + L <sub>1,2</sub> -GRMF [77]	–	–	–	–
RLS-WNN [25]	0.7755 (0.0093)	0.7669 (0.0140)	0.8524 (0.0072)	0.8390 (0.0261)
CMF [35]	0.7952 (0.0110)	0.7576 (0.0125)	0.8067 (0.0067)	0.8124 (0.0228)
NRLMF [36]	0.871 (0.017)	0.813 (0.027)	0.895 (0.011)	0.900 (0.021)
TMF [28]	0.843 (0.012)	0.819 (0.011)	0.882 (0.009)	0.886 (0.017)
DDR [80]	0.84	0.94	0.91	0.90
Proposed method	<b>0.9518 (0.0033)</b>	0.9682 (0.0078)	<b>0.9600 (0.0075)</b>	<b>0.9663 (0.105)</b>

MGRNNM has time complexity in each iteration  $O(n^3 + n^b \log(n))$ . The proposed method uses the RRSSVD\_DTI algorithm to solve NNM. According to Hastie et al. study [68], this algorithm requires  $O(r^3)$  time per each iteration. It can be said that the time complexity of the proposed algorithm in each iteration is equal to  $O(r^3 + n^b \log(n))$  and since  $r \ll \min(m, n)$ , the time complexity of the proposed algorithm is minor than MGRNNM.

In the following, more details of the time complexity are presented. The proposed method and MGRNNM are compared on four datasets based on running time. As shown in Fig. 3, the proposed method performed well on all the datasets.

It should be noted that in these applications, the online web detection system is designed to meet the needs of many people simultaneously. In such scenarios,

**Table 6** AUPR results for interaction prediction under validation setting CVS3

Method	Datasets			
	E	IC	GPCR	NR
MGRNNM [11]	0.9041 (0.0125)	0.9029 (0.0024)	0.7228 (0.0323)	0.5418 (0.0309)
SMGRNNM [11]	0.8087 (0.0156)	0.8079 (0.0096)	0.5963 (0.0336)	0.4356 (0.0177)
MC[78]	0.0124 (0.0005)	0.0421 (0.0043)	0.0549 (0.0105)	0.0850 (0.0227)
MCG[79]	0.0691 (0.0009)	0.2256 (0.0038)	0.1061 (0.0027)	0.2669 (0.0288)
WGRMF [20]	0.8070 (0.0185)	0.8128 (0.0069)	0.6093 (0.0314)	0.4643 (0.0183)
WKNKN + L <sub>1,2</sub> -GRMF [77]	0.799 (0.016)	0.826(0.008)	0.617 (0.024)	0.519 (0.038)
RLS-WNN[25]	0.5465 (0.0144)	0.7437 (0.0088)	0.5397 (0.0193)	0.4907 (0.0326)
CMF [35]	0.7808 (0.0131)	0.7786 (0.0108)	0.5989 (0.0323)	0.4774 (0.0173)
NRLMF [36]	0.812 (0.018)	0.785 (0.028)	0.556 (0.038)	0.449 (0.079)
TMF [28]	0.866 (0.007)	0.853 (0.008)	0.677 (0.028)	<b>0.675 (0.062)</b>
DDR [80]	0.82	0.80	0.61	0.64
Proposed Method	<b>0.9063 (0.0120)</b>	<b>0.9006 (0.0040)</b>	<b>0.7306 (0.0254)</b>	0.5496 (0.0307)

**Table 7** AUC results for interaction prediction under validation setting CVS3

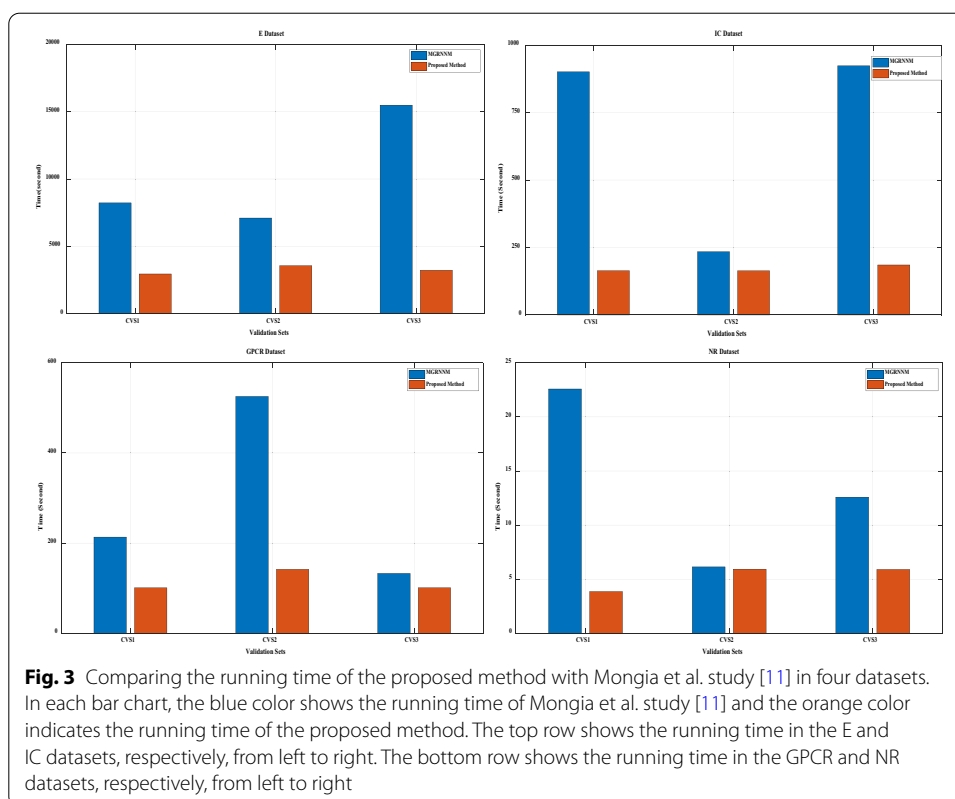
Method	Datasets			
	E	IC	GPCR	NR
MGRNNM [11]	0.9683 (0.0043)	0.9541 (0.0019)	0.8975 (0.0093)	0.7502 (0.0285)
SMGRNNM [11]	0.9246 (0.0091)	0.9346 (0.0041)	0.8798 (0.0134)	0.7263 (0.0211)
MC[78]	0.5234 (0.0057)	0.4724 (0.0065)	0.5683 (0.0310)	0.3767 (0.0204)
MCG[79]	0.8065 (0.0012)	0.7871 (0.0069)	0.6289 (0.0151)	0.6522 (0.0297)
WGRMF [20]	0.9338 (0.0071)	0.9460 (0.0034)	0.8892 (0.0110)	0.7967 (0.0132)
WKNKN + L <sub>1,2</sub> -GRMF [77]	–	–	–	–
RLS-WNN[25]	0.9067 (0.0105)	0.9286 (0.0046)	0.8694 (0.0146)	0.8124 (0.0202)
CMF [35]	0.9272 (0.0050)	0.9368 (0.0032)	0.8966 (0.0073)	0.8373 (0.0083)
NRLMF [36]	0.966 (0.005)	0.964 (0.007)	0.930 (0.012)	0.851 (0.027)
TMF [28]	0.976 (0.001)	0.972 (0.002)	<b>0.959 (0.007)</b>	<b>0.929 (0.029)</b>
DDR [80]	0.92	0.97	0.93	0.88
Proposed method	<b>0.9770 (0.0038)</b>	<b>0.9830 (0.0013)</b>	0.9087 (0.0075)	0.7694 (0.0249)

response time can be very significant, which shows the proper performance of the proposed method in time complexity that our approach can be used in DTI.

## Discussion

The use of adjacency matrices to represent the interaction between drug and target has been considered by many researchers in recent years. In DTI prediction, the detection of low-rank interaction has been significant. Using LRMA methods to solve these problems can improve performance. The results reported in this article showed that unifying matrix factorization and nuclear norm minimization approaches based on similarity matrix has a good effect on solving low-rank problem.

AUC, AUPR and running time are commonly used to evaluate the performance of methods in DTI. It should be noted that the use of a graph to express the adjacency of



the target and the drug in this application has improved the performance of the proposed method. Overall, this paper presents a powerful and fast method for applying DTI, which shows improved performance in four benchmark datasets.

To evaluate the proposed method, all four datasets are divided into three cross-validation settings. Results were reported as mean and variance of AUC and AUPR per 5 runs. It should be noted that the proposed method is compared with several methods [11, 20, 25, 28, 35, 36, 78–80]. The results have been shown the proposed method which was based on the similarity matrix, had the best performance. The time complexity of the proposed method was more appropriate than other methods.

## Conclusion

In drug-related processes such as drug discovery, drug side-effect prediction and drug repurposing, the interaction between drugs and targets (proteins) is very important. Drugs effect on targets (proteins) by altering the pharmaceutical functions of targets, such as enzymes, ion channels, G protein-coupled receptors (GPCRs), and nuclear receptors. DTIs analysis requires costly and time-consuming experiments. In this regard, CP-based approaches have been used to narrow down the search space and also reduce the cost and time of experiments.

In this paper, CP based on chemogenomic methods in DTI is used. It was shown that the use of similarity matrices in this application provides the best performance compared to the other methods. Also, we presented that the use of unifying of graph

regularized nuclear norm with bilinear factorization can be very effective in predicting DTI. In this paper, the proposed method on four datasets based on three different cross-validation settings is compared with six state-of-the-art methods. The results show a better performance of the proposed method. In general, the superiority of the proposed method can be expressed as follows:

- There is a trade-off parameter between the loss function and the low-rank regularization as  $\lambda$  in the NNM approach which is induced by the nuclear norm. The use of low-rank priors to many applications has been developed by these models where  $Z$  (in Eq. (3)) is low rank but its rank is not known a priori. These models also have multiple problems despite their convexity and theoretical guidelines for the choice of  $\lambda$  [9]. First, it is unclear how to impose a certain rank in  $Z$ : adjusting  $\lambda$  so  $Z$  has a predetermined rank usually produces unpleasant results than imposing it directly in Eq. (3) in many works. Second, it is impossible to obtain the  $Z$  factorization in Eq. (7) Causes not use the “kernel trick”. Third, Eq. (7) is a Semidefinite Program (SDP). Off-the-shelf SDP optimizers just divide into hundreds of variables, not amenable to the high dimensionality typically found in DTI problems. While many studies improve this issue, they still perform an SVD of  $Z$  in each iteration and make them disproportionate for managing dense and large-scale datasets. This paper indicates many nuclear norms regularized problems of the form (7) which can be optimized with a bilinear factorization of  $Z = UV^T$  by using the variational definition of the nuclear norm. In this paper, a unification of traditional bilinear factorization and nuclear norm approaches under one formulation in DTI applications have been proposed. Based on this result, we can analyze the conditions that both methods are equal and offer the best solution when they are not. This article explains how the proposed method can be used in DTI application. In the reference [9], the method based on nuclear norm regularization has been used. The optimization equations of the proposed method and its solution are expressed differently.
- Unifying nuclear norm with bilinear factorization is presented based on the similarity of drug-drug and target-target, which has caused the advantages of both methods to be combined.
- Rank-Restricted Soft Singular Value Decomposition method is used to optimize the nuclear norm minimization in the DTI problem. This method has not been used in this application so far. It was shown that the use of this method could have appropriate performance in data based on graph similarity.
- One of the critical parameters in evaluating the performance of these methods is the running time. With the increasing growth of this data, the use of computational methods will increase; on the other hand more similarity measures and samples with more features can be used. The time complexity of the running time is  $O(r^3 + n^b \log(n))$  for each iteration in the proposed algorithm. This complexity is polynomial, which has performed better than other new methods.
- The performance of the proposed method based on the AUC and AUPR measures had the best performance and also the results were suitable in datasets with different features.

- The proposed method can be widely used in other applications of CP. Using the proposed method does not require a complex process.

### Abbreviations

DTIs: Drug–target interactions; RRSSVD: Rank-Restricted Soft Singular Value Decomposition; E: Enzymes; ICs: Ion channels; GPCRs: G protein-coupled receptors; NRs: Nuclear receptors; AUC: Area Under the ROC Curve; AUPR: Area Under the Precision-Recall curve; CP: Computational prediction; NNM: Nuclear norm minimization; SVs: Singular values; LRMF: Low rank matrix factorization; SVD: Singular Value Decomposition; LS: Least Squares; NNP: Nuclear norm proximal; CVS: Cross-validation settings; CV: Cross-validation; WGRMF: Weighted Graph Regularized Matrix Factorization; CMF: Matrix Factorization; MSCMF: Multiple Similarities Collaborative Matrix Factorization; RLS-WNN: Regularized Least square nearest neighbor profile; MGRNNM: Multi Graph Regularized Nuclear Norm Minimization; MC: Matrix completion; MCG: Matrix completion on graphs, NRLMF: Neighborhood Regularized Logistic Matrix Factorization.

### Acknowledgements

Not applicable.

### Authors' contributions

AGS and ZA designed the novel method. AGS and MI implemented the proposed method. AGS, MI and JP performed interpretation on result of implementation. AGS and ZA wrote the manuscript. All authors read and approved the final manuscript.

### Funding

Not applicable.

### Availability of data and materials

Datasets are simulated from public databases which at the link: <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/> are publicly available. The authors declare that they have provided the code and data publicly accessible, which can be downloaded from: <https://github.com/ali289/DTI-GRNNwBF/>

### Declarations

#### Ethics approval and consent to participate

No ethics approval was required for the study.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Faculty of Electrical and Computer Engineering, University of Science and Technology of Mazandaran, P.O. Box, 48518-78195 Behshahr, Iran. <sup>2</sup>School of Medicine, Faculty of Medical Biotechnology, Shahrood University of Medical Sciences, Shahrood, Iran. <sup>3</sup>Faculty of Computer Engineering and IT, Payam Noor University, Tehran, Iran.

Received: 26 August 2020 Accepted: 29 October 2021

Published online: 17 November 2021

### References

1. Sachdev K, Gupta MK. A comprehensive review of feature based methods for drug target interaction prediction. *J Biomed Inform.* 2019;93:103159.
2. Mongia A, Jain V, Chouzenoux E, Majumdar A. Deep latent factor model for predicting drug target interactions. In: *ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2019; p. 1254–8.
3. K. Sachdev, M.K. Gupta, A hybrid ensemble-based technique for predicting drug–target interactions. *Chem Biol Drug Des.*, 2020.
4. Deng Y., Xu X, Qiu Y, Xia J, Zhang W, Liu S. A multimodal deep learning framework for predicting drug-drug interaction events. *Bioinformatics*, 2020.
5. Fakhraei S, Huang B, Raschid L, Getoor L. Network-based drug–target interaction prediction with probabilistic soft logic. *IEEE/ACM Trans Comput Biol Bioinform.* 2014;11:775–87.
6. Ezzat A, Wu M, Li X-L, Kwok C-K. Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinform.* 2019;20:1337–57.

7. Yin M-M, Liu J-X, Gao Y-L, Kong X-Z, Zheng C-H. NCPLP: a novel approach for predicting microbe-associated diseases with network consistency projection and label propagation. *IEEE Trans Cybern.* 2020.
8. Cui Z, Liu J-X, Gao Y-L, Zhu R, Yuan S-S. LncRNA-disease associations prediction using bipartite local model with nearest profile-based association inferring. *IEEE J Biomed Health Inform.* 2019;24:1519–27.
9. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol.* 2007;25:197–206.
10. Hendrickson JB. Concepts and applications of molecular similarity. *Science.* 1991;252:1189–90.
11. Mongia A, Majumdar A. Drug–target interaction prediction using multi graph regularized nuclear norm minimization. *Plos One.* 2020;15:e0226484.
12. Li H, Gao Z, Kang L, Zhang H, Yang K, Yu K, et al. TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.* 2006;34:W219–24.
13. Yildirim MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M. Drug–target network. *Nat Biotechnol.* 2007;25:1119–26.
14. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics.* 2008;24:i232–40.
15. Mousavian Z, Masoudi-Nejad A. Drug–target interaction prediction via chemogenomic space: learning-based methods. *Expert Opin Drug Metab Toxicol.* 2014;10:1273–87.
16. Ezzat A, Wu M, Li X-L, Kwoh C-K. Drug–target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinform.* 2016;17:267–76.
17. Yu H, Chen J, Xu X, Li Y, Zhao H, Fang Y, et al. A systematic prediction of multiple drug–target interactions from chemical, genomic, and pharmacological data. *PLoS One.* 2012;7:e37608.
18. He Z, Zhang J, Shi X-H, Hu L-L, Kong X, Cai Y-D, et al. Predicting drug–target interaction networks based on functional groups and biological features. *PLoS One.* 2010;5:e9603.
19. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug–target interactions and drug repositioning via network-based inference. *PLoS Comput Biol.* 2012;8:e1002503.
20. Ezzat A, Zhao P, Wu M, Li X-L, Kwoh C-K. Drug–target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Trans Comput Biol Bioinform.* 2016;14:646–56.
21. Thafar MA, Olayan RS, Ashoor H, Albaradei S, Bajic VB, Gao X, et al. DTIGEMS+: drug–target interaction prediction using graph embedding, graph mining, and similarity-based techniques. *J Cheminform.* 2020;12:1–17.
22. Ding H, Takigawa I, Mamitsuka H, Zhu S. Similarity-based machine learning methods for predicting drug–target interactions: a brief review. *Brief Bioinform.* 2014;15:734–47.
23. Bleakley K, Yamanishi Y. Supervised prediction of drug–target interactions using bipartite local models. *Bioinformatics.* 2009;25:2397–403.
24. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics.* 2011;27:3036–43.
25. Van Laarhoven T, Marchiori E. Predicting drug–target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS One.* 2013;8:e66952.
26. Shi J-Y, Liu Z, Yu H, Li Y-J. Predicting drug–target interactions via within-score and between-score. *BioMed Res Int.* 2015;20:15.
27. Shi J-Y, Yiu S-M, Li Y, Leung HC, Chin FY. Predicting drug–target interaction for new drugs using enhanced similarity measures and super-target clustering. *Methods.* 2015;83:98–104.
28. Shi J-Y, Zhang A-Q, Zhang S-W, Mao K-T, Yiu S-M. A unified solution for different scenarios of predicting drug–target interactions via triple matrix factorization. *BMC Syst Biol.* 2018;12:45–56.
29. Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics.* 2020;36:5545–7.
30. Monteiro NR, Ribeiro B, Arrais J. Drug–target interaction prediction: end-to-end deep learning approach. *IEEE/ACM Trans Comput Biol Bioinform.* 2020.
31. Wang Y-B, You Z-H, Yang S, Yi H-C, Chen Z-H, Zheng K. A deep learning-based method for drug–target interaction prediction based on long short-term memory neural network. *BMC Med Inform Decis Mak.* 2020;20:1–9.
32. Ay M, Goh K-I, Cusick ME, Barabasi A-L, Vidal M. Drug–target network. *Nat Biotechnol.* 2007;25:1119–27.
33. Seal A, Ahn Y-Y, Wild DJ. Optimizing drug–target interaction prediction based on random walk on heterogeneous networks. *J Cheminform.* 2015;7:1–12.
34. Gönen M. Predicting drug–target interactions from chemical and genomic kernels using Bayesian matrix factorization. *Bioinformatics.* 2012;28:2304–10.
35. Zheng X, Ding H, Mamitsuka H, Zhu S. Collaborative matrix factorization with multiple similarities for predicting drug–target interactions. In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 2013*, pp. 1025–1033.
36. Liu Y, Wu M, Miao C, Zhao P, Li X-L. Neighborhood regularized logistic matrix factorization for drug–target interaction prediction. *PLoS Comput Biol.* 2016;12:e1004760.
37. Cui Z, Gao Y-L, Liu J-X, Wang J, Shang J, Dai L-Y. The computational prediction of drug-disease interactions using the dual-network L<sub>2,1</sub>-CMF method. *BMC Bioinform.* 2019;20:1–10.
38. Zhang Z-C, Zhang X-F, Wu M, Ou-Yang L, Zhao X-M, Li X-L. A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks. *Bioinformatics.* 2020;36:3474–81.
39. Gu S, Xie Q, Meng D, Zuo W, Feng X, Zhang L. Weighted nuclear norm minimization and its applications to low level vision. *Int J Comput Vis.* 2017;121:183–208.
40. Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? *J ACM (JACM).* 2011;58:1–37.
41. Song G-J, Ng MK. Nonnegative low rank matrix approximation for nonnegative matrices. *Appl Math Lett.* 2020;105:106300.

42. Xia S, Song J, Chen D, Wang J. Uncertainty quantification for hyperspectral image denoising frameworks based on low-rank matrix approximation. *arXiv preprint arXiv:2004.10959*, 2020.
43. Eriksson A, Van Den Hengel A. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the L 1 norm. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011; p. 771–778.
44. Chi Y, Lu YM, Chen Y. Nonconvex optimization meets low-rank matrix factorization: an overview. *IEEE Trans Signal Process.* 2019;67:5239–69.
45. Zhou D, Cao Y, Gu Q. Accelerated factored gradient descent for low-rank matrix factorization. In: International conference on artificial intelligence and statistics, 2020, p. 4430–40.
46. Huang Z, Salama P, Shao W, Zhang J, Huang K. Low-rank reorganization via proportional hazards non-negative matrix factorization unveils survival associated gene clusters, *arXiv preprint arXiv:2008.03776*, 2020.
47. Lei B, Cheng N, Frangi AF, Tan E-L, Cao J, Yang P, et al. Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease. *Med Image Anal.* 2020;61:101652.
48. Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Brief Bioinform.* 2020.
49. Brzyski D, Hu X, Goni J, Ances B, Randolph TW, Harezlak J. A sparsity inducing nuclear-norm estimator (SpINNER) for matrix-variate regression in brain connectivity analysis, *arXiv preprint arXiv:2001.11548*, 2020.
50. Srebro N, Jaakkola T. Weighted low-rank approximations. In: Proceedings of the 20th international conference on machine learning (ICML-03), 2003; p. 720–7.
51. Zhao Z, Wang S, Wong D, Guo Y, Chen X. The sparse and low-rank interpretation of SVD-based denoising for vibration signals. In: 2020 IEEE international instrumentation and measurement technology conference (I2MTC), 2020, pp. 1–6.
52. Xu S, Zhang C, Zhang J. Adaptive quantile low-rank matrix factorization. *Pattern Recognit.* p. 107310; 2020.
53. Zhao Q, Meng D, Xu Z, Zuo W, Yan Y.  $\$ L_1$   $\$$ -norm low-rank matrix factorization by variational Bayesian method. *IEEE Trans Neural Netw Learn Syst.* 2015;26:825–39.
54. Kong Y, Shao M, Li K, Fu Y. Probabilistic low-rank multitask learning. *IEEE Trans Neural Netw Learn Syst.* 2017;29:670–80.
55. Tu W, Liu P, Zhao J, Liu Y, Kong L, Li G, et al. M-estimation in low-rank matrix factorization: a general framework. In: 2019 IEEE international conference on data mining (ICDM), 2019; pp. 568–77.
56. Baldi P, Hornik K. Neural networks and principal component analysis: learning from examples without local minima. *Neural Netw.* 1989;2:53–8.
57. Gillis N, Glineur F. Low-rank matrix approximation with weights or missing data is NP-hard. *SIAM J Matrix Anal Appl.* 2011;32:1149–65.
58. Jain P, Oh S. Provable tensor factorization with missing data. In: Advances in Neural Information Processing Systems, 2014, p. 1431–9.
59. Hladik M, Hartman D, Zamani M. Maximization of a PSD quadratic form and factorization. *Optim Lett.*, 2020;pp. 1–14.
60. Gu S, Zhang L, Zuo W, Feng X. Weighted nuclear norm minimization with application to image denoising. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014; p. 2862–9.
61. Yang Y, Yang Z, Li J, Fan L. Foreground-background separation via generalized nuclear norm and structured sparse norm based low-rank and sparse decomposition. *IEEE Access.* 2020;8:84217–29.
62. Jiang T-X, Huang T-Z, Zhao X-L, Deng L-J. Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm. *J Comput Appl Math.* 2020;372:112680.
63. Zhu W, Peng B. Sparse and low-rank regularized deep subspace clustering. *Knowl-Based Syst.* 2020; p. 106199.
64. Sun X, Wang Y, Zhang X. Multi-view subspace clustering via non-convex tensor rank minimization. In: 2020 IEEE international conference on multimedia and expo (ICME), 2020, p. 1–6.
65. Li J, Fan W, Li Y, Qian Z. Low-frequency noise suppression in desert seismic data based on an improved weighted nuclear norm minimization algorithm. *IEEE Geosci Remote Sens Lett.* 2020.
66. Huang D, Cabral R, De la Torre F. Robust regression. *IEEE Trans Pattern Anal Mach Intell.* 2015;38:363–75.
67. Cabral R, De la Torre F, Costeira JP, Bernardino A. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In: Proceedings of the IEEE international conference on computer vision, 2013; p. 2488–95.
68. Hastie T, Mazumder R, Lee JD, Zadeh R. Matrix completion and low-rank SVD via fast alternating least squares. *J Mach Learn Res.* 2015;16:3367–402.
69. Recht B, Fazel M, Parrilo PA. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* 2010;52:471–501.
70. Hattori M, Tanaka N, Kanehisa M, Goto S. SIMCOMP/SUBCOMP: chemical structure search servers for network analyses. *Nucleic Acids Res.* 2010;38:W652–6.
71. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 1981;147:195–7.
72. May JP. Equivariant homotopy and cohomology theory, vol. 91 of CBMS Regional Conference Series in Mathematics. In: Published for the conference board of the mathematical sciences, Washington, DC, 1996, p. 88
73. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 2006;34:D354–7.
74. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, et al. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* 2004;32:D431–3.
75. Günther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, et al. SuperTarget and Matador: resources for exploring drug–target relationships. *Nucleic Acids Res.* 2007;36:D919–22.
76. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008;36:D901–6.



77. Cui Z, Gao Y-L, Liu J-X, Dai L-Y, Yuan S-S. L<sub>2, 1</sub>-GRMF: an improved graph regularized matrix factorization method to predict drug–target interactions. *BMC Bioinform.* 2019;20:1–13.
78. Majumdar A, Ward RK. Some empirical advances in matrix completion. *Signal Process.* 2011;91:1334–8.
79. Kalofolias V, Bresson X, Bronstein M, Vandergheynst P. Matrix completion on graphs. arXiv preprint arXiv:1408.1717, 2014.
80. Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics.* 2018;34:1164–73.
81. Kiriinnis P. Fast algorithms for the Sylvester equation  $AX - XBT = C$ . *Theoret Comput Sci.* 2001;259:623–38.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

