




Systems biology

# deepDR: a network-based deep learning approach to *in silico* drug repositioning

Xiangxiang Zeng<sup>1,†</sup>, Siyi Zhu <sup>1,†</sup>, Xiangrong Liu <sup>1</sup>, Yadi Zhou<sup>2</sup>,  
Ruth Nussinov<sup>3,4</sup> and Feixiong Cheng <sup>2,5,6,\*</sup>

<sup>1</sup>Department of Computer Science, Xiamen University, Xiamen 361005, China, <sup>2</sup>Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA, <sup>3</sup>Computational Structural Biology Section, Basic Science Program, Frederick National Laboratory for Cancer Research, National Cancer Institute at Frederick, Frederick, MD 21702, USA, <sup>4</sup>Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel, <sup>5</sup>Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA and <sup>6</sup>Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

\*To whom correspondence should be addressed.

<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate editor: Lenore Cowen

Received on January 22, 2019; revised on April 22, 2019; editorial decision on May 10, 2019; accepted on May 14, 2019

## Abstract

**Motivation:** Traditional drug discovery and development are often time-consuming and high risk. Repurposing/repositioning of approved drugs offers a relatively low-cost and high-efficiency approach toward rapid development of efficacious treatments. The emergence of large-scale, heterogeneous biological networks has offered unprecedented opportunities for developing *in silico* drug repositioning approaches. However, capturing highly non-linear, heterogeneous network structures by most existing approaches for drug repositioning has been challenging.

**Results:** In this study, we developed a network-based deep-learning approach, termed deepDR, for *in silico* drug repurposing by integrating 10 networks: one drug–disease, one drug–side-effect, one drug–target and seven drug–drug networks. Specifically, deepDR learns high-level features of drugs from the heterogeneous networks by a multi-modal deep autoencoder. Then the learned low-dimensional representation of drugs together with clinically reported drug–disease pairs are encoded and decoded collectively via a variational autoencoder to infer candidates for approved drugs for which they were not originally approved. We found that deepDR revealed high performance [the area under receiver operating characteristic curve (AUROC) = 0.908], outperforming conventional network-based or machine learning-based approaches. Importantly, deepDR-predicted drug–disease associations were validated by the ClinicalTrials.gov database (AUROC = 0.826) and we showcased several novel deepDR-predicted approved drugs for Alzheimer’s disease (e.g. risperidone and aripiprazole) and Parkinson’s disease (e.g. methylphenidate and pergolide).

**Availability and implementation:** Source code and data can be downloaded from <https://github.com/ChengF-Lab/deepDR>

**Contact:** [chengf@ccf.org](mailto:chengf@ccf.org)

**Supplementary information:** [Supplementary data](#) are available online at *Bioinformatics*.

## 1 Introduction

Medical researchers have long sought to uncover single molecule defects that cause human diseases with the goal of developing ‘magic-bullet’-targeted therapies. However, this ‘one gene, one drug, one disease’ reductionism-informed paradigm overlooks the inherent complexity of diseases and continue to challenge personalized diagnosis and drug discovery (Greene and Loscalzo, 2017). A recent study estimated that it costed pharmaceutical companies \$2.6 billion in 2015 to develop a new US Food and Drug Administration (FDA)-approved drug, as compared to \$802 million in 2003 (Avorn, 2015). While there are many factors that contribute to this limited approval rate, one important, often overlooked determinant is the continued adherence to the classical ‘one gene, one drug, one disease’ hypothesis of drug development dating to the work of Ehrlich (Tan and Grimes, 2010). As drug targets do not operate in isolation from the complex system of proteins that comprise the molecular machinery of the cell with which they associate, we believe that each drug–target interaction must be examined in an appropriate integrative context (Cheng et al., 2019; Greene and Loscalzo, 2017). Doing so will offer novel insights into drug mechanisms, potential adverse effects and so-called off-target effects that may be used to repurpose drugs rationally, termed drug repurposing/repositioning (Cheng et al., 2018; Pushpakom et al., 2018). However, drug repurposing is fraught with challenges due to the unknown underlying complex pharmacology and biology.

Among the advances, capable and intelligent computer science-based algorithms offer an unbiased, rational roadmap for drug repositioning in multiple complex diseases for which they were not originally approved, by integrating large-scale genomic and phenotypic data, as well as the chemical and bioactivity data from hundreds of approved drugs (Cheng et al., 2018; Pushpakom et al., 2018). For example, heterogeneous data sources and various drug or disease similarity networks provide diverse information and a multi-layer perspective for predicting novel drug–disease associations. Therefore, incorporating multiple data sources can potentially boost the accuracy of *in silico* drug repositioning (Ching et al., 2018). However, most existing methods for drug repositioning are limited to only drug similarity networks, disease similarity network or bipartite drug–disease models (Cheng et al., 2012, 2017). Furthermore, those approaches cannot be directly extended to take into account the heterogeneous nodes or network topological information among different biological networks (Cheng et al., 2018; Pushpakom et al., 2018).

Informative network-based features play essential roles in the prediction of drug–disease relationships. Yet, there are several challenges to learn an informative and low-dimensional network representation (also known as *network embedding*), while preserving the network structures from heterogeneous data sources of drug-related networks. In particular, drugs with the same or similar functional annotations in these networks often exhibit a complex mixture of relationships, based on both homophily (close proximity to each other in the network) and structural similarity (similar structural roles, regardless of the position in the network). Thus, it is a challenging task to learn a low-dimensional embedding of drugs (nodes) that preserves non-linear network structure while remaining predictive for novel drug indications. Even more challenging is the construction of such a compact low-dimensional embedding of drugs that is consistent across different drug functional and molecular interaction modalities, such as across different types of drug-related networks (Glorigjevic et al., 2018).

Deep learning is a promising technique for capturing complex and highly non-linear network structure, offering powerful tools for many scenarios, such as speech recognition, image classification, and natural

language processing, as well as for medicine and biology (Angermueller et al., 2016; Ching et al., 2018; Topol, 2019). In this study, we developed a new approach, termed deepDR (deep learning-based drug repositioning), to systematically infer new drug–disease relationships for *in silico* drug repurposing. The underlying concept of deepDR is to fuse diverse information from different types of networks and infer new applications for existing drugs that were not originally approved by a collective variational autoencoder (cVAE). The advantages of deepDR can be summarized as follows: (i) deepDR integrates diverse information from nine heterogeneous networks, which can potentially boost the accuracy of drug–disease prediction and offer new insights into drug repositioning; (ii) deepDR preserves the non-linear network structure by applying multiple layers of non-linear functions, which is able to capture complex topological patterns across multiple types of networks; (iii) deepDR uses side information (drug features) to pre-train a variational autoencoder, which overcomes drug–disease rating sparsity by feeding both ratings and side information into the same inference network and generation network. We found that deepDR revealed high performance in predicting drug–disease associations from the clinically reported drug–disease network, outperforming previous state-of-the-art approaches. Importantly, we demonstrated that deepDR had high accuracy on an external validation set collected from the ClinicalTrials.gov database, suggesting a high generalization ability.

## 2 Materials and methods

### 2.1 Re-construction of heterogeneous networks

We assembled clinically reported or experimentally validated drug–disease network by assembling data from two commonly used databases: DrugBank (Wishart et al., 2018) and repoDB (Brown and Patel, 2017). Chemical name, generic name or commercial name of each drug were standardized by Medical Subject Headings (MeSH) and Unified Medical Language System (UMLS) vocabularies (Bodenreider, 2004) and further converted to DrugBank ID from the DrugBank database (v4.3) (Law et al., 2014). Generic name for each disease was annotated by MeSH. In total, we constructed nine networks: (i) clinically reported drug–drug interactions, (ii) drug–target interactions, (iii) drug–side-effect associations, (iv) chemical similarities, (v) therapeutic similarities derived from the Anatomical Therapeutic Chemical Classification System, (vi) drugs’ target sequence similarities, (vii) Gene Ontology (GO) *biological process*, (viii) GO *cellular component* and (ix) GO *molecular function*. More details for building heterogeneous networks for drugs are provided in the [Supplementary Material](#). In addition, 6677 clinically reported drug–disease pairs connecting 1519 drugs and 1229 diseases were further collected for building predictive deep learning models ([Supplementary Tables S1 and S2](#)). For the external validation set, we assembled the most recent drug–disease associations from the ClinicalTrials.gov database (<https://clinicaltrials.gov/>), by excluding existing pairs in the abovementioned DrugBank (Wishart et al., 2018) and repoDB (Brown and Patel, 2017) databases.

### 2.2 Random walk-based network representation

We adopt the approach used in Cao (2016) to capture network structural information and to characterize the topological context of each drug. The vertices of a network are first ordered randomly. Assuming the current vertex is the *i*th vertex, a transition matrix *A* captures the transition probabilities between different vertices. It takes both local and global topological connectivity patterns within the network into consideration to fully exploit the underlying direct or

indirect relationships between nodes. Thus, at each time, the random walk procedure will continue with a probability of  $\omega$ , and will return to the original vertex and restart the procedure with a probability  $1 - \omega$ . This step leads to the recurrence relation as follow:

$$p_k = \omega \cdot p_{k-1}A + (1 - \omega)p_0 \quad (1)$$

Where  $p_k$  is a row vector, whose  $j$ th entry indicates the probability of reaching the  $j$ th vertex after  $k$  steps of transitions and  $p_0$  is the initial 1-hot vector with the value of the  $i$ th entry being 1 and all other entries being 0. By summing each random walk of  $p_k$  and repeating the process for each node in the network, we can get a probabilistic co-occurrence matrix.

Next, we calculated a shifted *positive pointwise mutual information* (PPMI) matrix (Bullinaria and Levy, 2007). The PPMI matrix can be viewed as a matrix factorization method which factorizes a co-occurrence matrix to yield network representations. The PPMI matrix can be constructed as follow:

$$\text{PPMI} = \max \left( \log \frac{M(i, j) * \sum_i \sum_j M(i, j)}{\sum_i M(i, j) * \sum_j M(i, j)}, 0 \right) \quad (2)$$

Where  $M$  is the original co-occurrence matrix,  $N_r$  is the number of rows,  $N_c$  is the number of columns. Negative PPMI values are changed to 0. The random walk-based representation mitigates the sparsity of some individual network types which acts as a pre-processing step prior to the deeper integration described in the next steps.

### 2.3 Network fusion via multi-modal deep autoencoder (MDA)

To get high-quality drug features that fuse multiple networks, we followed the strategy proposed previously (Gligorijevic et al., 2018), which integrates multiple networks represented by PPMI matrices using MDA. MDA constructs a low-dimensional feature representation of  $n$  drugs, that best approximates all networks, by projecting their PPMI matrices  $X^{(j)} \in \mathbb{R}^{n \times n}$  using multiple non-linear activation functions into a common feature space  $H_c \in \mathbb{R}^{d_c \times n}$ . An autoencoder is a special type of neural network that is composed of an encoding part and a decoding part (Vincent et al., 2010). We formulated the encoding and decoding part of the MDA in the following sections.

#### 2.3.1 Encoder

In the first hidden layer of the MDA, we first computed low-dimensional non-linear embedding  $H_{\text{encode}}^{(j)} \in \mathbb{R}^{d_j \times n}$ , for each network  $j \in \{1, \dots, N\}$ :

$$H_{\text{encode}}^{(j)} = \sigma(W_{\text{encode}}^{(j)} X^{(j)} + B_{\text{encode}}^{(j)})$$

Where  $W_{\text{encode}}^{(j)} \in \mathbb{R}^{d_j \times n}$  and  $B_{\text{encode}}^{(j)} \in \mathbb{R}^{d_j \times n}$  are weight and bias matrices.  $\sigma$  is the sigmoid activation function.

We then concatenated non-linear embeddings of all networks, and computed a common feature representation by applying multiple non-linear functions on them. There can be  $L$  layers after we get the common representation.

$$H_{c,1} = \sigma(W_1[H^{(1)}, \dots, H^{(N)}] + B_1) \\ H_{c,l+1} = \sigma(W_l H_{c,l} + B_l)$$

Where  $[H^{(1)}, \dots, H^{(N)}]$  are the concatenated activation matrices of  $N$  embeddings of the previous layers,  $l \in \{1, \dots, L\}$  is the layer number for the successive integrated embeddings.

#### 2.3.2 Decoder

We first computed the reconstructed common layer  $H_{c,2L}$ , with the same number of decoding layers as the encoding layers. Then we can compute individual representations for each network  $H_{\text{decode}}^{(j)}$ .

$$H_{\text{decode}}^{(j)} = \sigma(W_{\text{decode},1}^{(j)} H_{c,2L} + B_{\text{decode},1}^{(j)})$$

Later we computed the reconstructed PPMI matrices  $\hat{X}^{(j)}$  for each network.

$$\hat{X}^{(j)} = \sigma(W_{\text{decode},2}^{(j)} H_{\text{decode}}^{(j)} + B_{\text{decode},2}^{(j)})$$

The goal of MDA is to minimize the reconstruction loss between each original and reconstructed PPMI matrix, defined as follows:

$$\argmin_{\theta} \sum_{j=1}^N \text{Loss}(X^{(j)}, \hat{X}^{(j)}) \quad (3)$$

Where Loss is the sample-wise binary cross-entropy function and  $\theta = \{W_{\text{encode}}^{(j)}, B_{\text{encode}}^{(j)}, W_{\text{decode}}^{(j)}, B_{\text{decode}}^{(j)}, W_l, B_l\}$  for  $l \in \{1, \dots, 2L\}$  is the set of all parameters in both the encoding and decoding parts of the MDA to be learned in the training process. We chose the standard back-propagation algorithm to optimize the loss function.

### 2.4 Collective variational autoencoder

The extracted features from MDA serves as side information of drugs. We then used the cVAE model (Chen and De Rijke, 2018) to infer new drug-disease associations. cVAE encodes and decodes drug-disease associations and side information through the same inference network and generation network.

#### 2.4.1 Generation network

While drug-disease associations and drug features are two different types of information, cVAE assumes the output of the generation network to follow different distributions according to the type of input it has been fed. We defined drug-disease associations as  $Y$  and drug features as  $X$ . Following the common practice for VAE, we first assumed the latent variables  $u$  and  $z$  to follow a Gaussian distribution:

$$u \sim \mathcal{N}(0, I), z \sim \mathcal{N}(0, I)$$

where  $I \in \mathbb{R}^{k \times k}$  is an identity matrix and  $k$  is the dimension of latent drug representation. Motivated by the positive-unlabeled (PU) learning framework, where observed and unobserved entries are penalized differently in the objective (Elkan and Noto, 2008; Hsieh et al., 2014), we introduced a parameter  $\alpha$  to balance between positive samples and negative samples. While  $X$  and  $Y$  are fed into the same network, we would like to distinguish them via different distributions. For drug-disease associations, the rating of disease  $j$  over all drugs follows a Bernoulli distribution:

$$y_j | u_j \sim \text{Bernoulli}(\sigma(f_{\theta}(u_j)))$$

This defines the loss function when feeding drug-disease associations as input, i.e. the logistic log-likelihood for disease  $j$ :

$$\log p_{\theta}(y_j | u_j) = \sum_{i=1}^n \alpha \cdot y_{ji} \log \sigma(f_{ji}) + (1 - y_{ji}) \log (1 - \sigma(f_{ji})) \quad (4)$$

Where  $f_{ji}$  is the  $i$ th element of vector  $f_{\theta}(u_j)$ , and  $f_{\theta}(u_j)$  is normalized through a sigmoid function so that  $f_{ji}$  is within (0, 1).

For drug features, we focus on numerical features so that we assume the  $j$ th dimension of drug features follows a Gaussian distribution:

$$x_j|z_j \sim \mathcal{N}(f_\theta(z_j), I)$$

This defines the loss function when feeding drug features as input, i.e. the Gaussian log-likelihood for dimension  $j$ :

$$\log p_\theta(x_j|z_j) = \sum_{i=1}^n \frac{c_{ji}}{2} (x_{ji} - f_{ji})^2 \quad (5)$$

Where  $c_{ji} = \alpha$  if  $x_{ji} > 0$ , else  $c_{ji} = 1$ .  $f_{ji}$  is the  $i$ th element of vector  $f_\theta(z_j)$ . Note that although we assumed  $x$  and  $y$  to be generated from  $z$  and  $u$  respectively, the generation has shared parameters  $\theta$ .

### 2.4.2 Inference network

The log-likelihood of cVAE is intractable due to the non-linear transformations of the generation network. Thus, we resorted to variational inference to approximate the distribution. Variational inference approximates the true intractable posterior with a simpler variational distribution  $q(U, Z)$ . We followed the mean-field assumption (Xing et al., 2003) by setting  $q(U, Z)$  to be a fully factorized Gaussian distribution:

$$q(U, Z) = \prod_{j=1}^m q(u_j) \prod_{j=1}^d q(z_j)$$

$$q(u_j) \sim \mathcal{N}(\mu_j, \text{diag}(\sigma_j^2)), \quad q(z_j) \sim \mathcal{N}(\mu_{m+j}, \text{diag}(\sigma_{m+j}^2))$$

Besides, we replaced individual variational parameters with a data-dependent function by an inference network parameterized by  $\phi$ , i.e.  $f_\phi$ , where  $\mu_j$  and  $\sigma_j$  are generated as:

$$\mu_j = \mu(f_\phi(y_j)), \quad \sigma_j = \sigma(f_\phi(y_j)), \quad \forall j = 1, \dots, m$$

$$\mu_{m+j} = \mu(f_\phi(x_j)), \quad \sigma_{m+j} = \sigma(f_\phi(x_j)), \quad \forall j = 1, \dots, d$$

We can derive the evidence lower bound (ELBO):

$$\mathcal{L}(q) = \mathbb{E}_{q_\phi}[\log p_\theta(X, Y|U, Z)] - \beta \mathbb{KL}(q_\phi || p(U, Z)) \quad (6)$$

where  $\beta$  is a parameter introduced to control the strength of regularization. We used a Monte Carlo gradient estimator (Paisley et al., 2012) to infer the expectation. We drew  $L$  samples of  $u_j$  and  $z_j$  from  $q_\phi$  and performed stochastic gradient ascent to optimize the ELBO. In order to take gradients with respect to  $\phi$  through sampling, we followed the reparameterization risk (Kingma and Welling, 2013) to sample  $u_j$  and  $z_j$  as:

$$u_j^{(l)} = \mu(f_\phi(y_j)) + \epsilon_1^{(l)} \odot \sigma(f_\phi(y_j))$$

$$z_j^{(l)} = \mu(f_\phi(x_j)) + \epsilon_2^{(l)} \odot \sigma(f_\phi(x_j))$$

$$\epsilon_1^{(l)} \sim \mathcal{N}(0, I), \quad \epsilon_2^{(l)} \sim \mathcal{N}(0, I)$$

As the  $\mathbb{KL}$ -divergence can be analytically derived (Kingma and Welling, 2013), we can then rewrite  $\mathcal{L}(q)$  as:

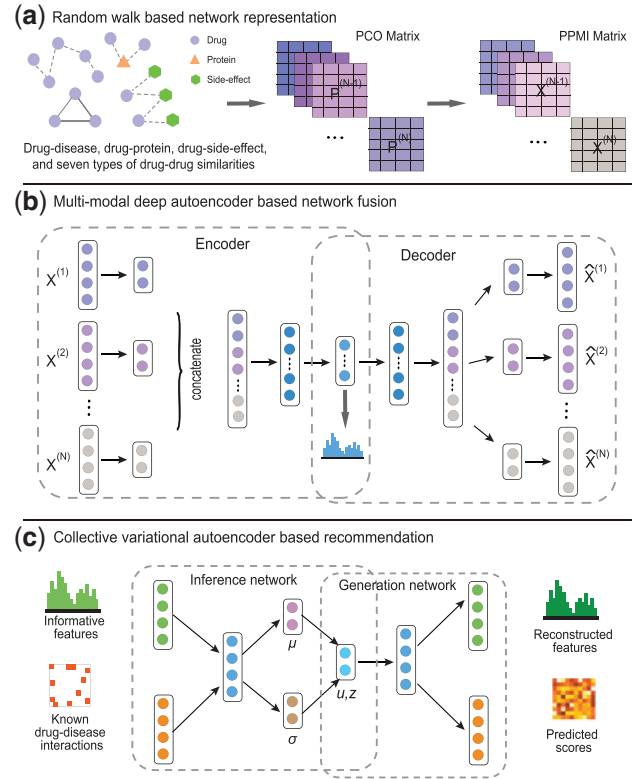
$$\mathcal{L}(q) = \frac{1}{L} \sum_{l=1}^L \left( \sum_{j=1}^m \log p_\theta(y_j|u_j^{(l)}) + \sum_{j=1}^d \log p_\theta(x_j|z_j^{(l)}) \right) + \sum_{j=1}^{d+m} \left( 1 + 2\log(\sigma_j) - \mu_j^2 - \sigma_j^2 \right) \quad (7)$$

We then maximized ELBO given above to learn  $\theta$  and  $\phi$ .

## 3 Results

### 3.1 Pipeline of deepDR

As shown in Figure 1, deepDR consists of three steps: (i) deepDR converts topological structure of each network into a high-quality vector representation by first applying the random walk with restart (RWR) method and then constructing a PPMI matrix capturing structural



**Fig. 1.** Pipeline of deepDR. (a) deepDR generates random walk-based network representation from a complicated heterogeneous network that contains 10 drug-related networks (see Section 2). (b) deepDR fuses PPMI (positive pointwise mutual information) matrices of each network into a compact, low-dimensional feature representation common to all networks via a multi-modal deep autoencoder (MDA). (c) deepDR uses a collective variational autoencoder (cVAE) to predict potential associations between drugs and diseases. Drug features and known (clinically reported or approved) drug-disease interactions are encoded and decoded collectively by the same inference network and generation network.

information of the network. (ii) deepDR fuses PPMI matrices of each network into a compact, low-dimensional feature representation common to all networks using MDA in an unsupervised way. In the early parts of the deep autoencoder, deepDR uses individual layers for handling each network type; later it connects all the layers into a single bottleneck layer. Consequently, we can extract high-quality features from this single bottleneck layer. (iii) deepDR uses cVAE to infer potential associations between drugs and diseases. Specifically, it feeds high-quality features extracted from the second step to the VAE for pre-training, and then refines the VAE by feeding the drug-disease association network.

### 3.2 Baseline methods

The prediction results are based on detailed comparisons between the models listed as follows. More details of the baseline methods and hyperparameter selection (Supplementary Tables S6–S9) can be found in the Supplementary Material.

- DTINet: The DTINet (Luo et al., 2017) focuses on learning a low-dimensional vector representation of features from the heterogeneous network and then applies inductive matrix completion (IMC) (Nagarajan and Dhillon, 2014) to make predictions based on the learned representations.



- **KBMF**: Kernelized Bayesian matrix factorization method (Gönen *et al.*, 2013) can make use of multiple side information sources and can be applied in recommender systems.
- **RF**: Random forest (Breiman, 2001) represents a collection of decision trees, which are grown from bootstrap samples of the training data without pruning, and make predictions based on majority votes of the ensemble trees.
- **SVM**: Support vector machine (Cortes and Vapnik, 1995) is based on a statistical learning theory derived from the structural risk minimization principle and Vapnik–Chervonenkis (VC) dimension.
- **RWR**: Random walk with restart (Cao *et al.*, 2014; Köhler *et al.*, 2008), a network diffusion algorithm, which is useful in measuring the proximity between two nodes of a network.
- **Katz**: The Katz (Singh-Blom *et al.*, 2013) measure is a graph-based method for finding node similarity to a given node by computing how many different path lengths exist between the pair.

### 3.3 Performance of deepDR on the cross-validation

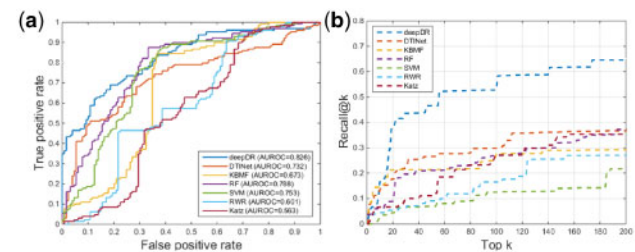
To evaluate the performance of deepDR, we first performed 5-fold cross-validation. In total, we assembled 6677 clinically reported drug–disease pairs connecting 1519 approved drugs and 1229 human disease terms. During the 5-fold cross-validation, we randomly selected a subset of 20% of the clinically reported drug–disease pairs and a matching number of randomly sampled unknown pairs as the test set, and the remaining 80% clinically reported drug–disease pairs with same number of randomly sampled unknown pairs were used to train the model. The area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPR) were utilized to evaluate the overall performance of deepDR. To reduce the data bias of cross-validation, it was repeated 10 times and the average performance was computed. We found that deepDR showed high accuracy (AUROC = 0.908 and AUPR = 0.923) in 5-fold cross validation, outperforming the state-of-the-art methods: DTINet (AUROC = 0.862 and AUPR = 0.892), KBMF (AUROC = 0.791 and AUPR = 0.826), RF (AUROC = 0.783 and AUPR = 0.805), SVM (AUROC = 0.771 and AUPR = 0.778), RWR (AUROC = 0.708 and AUPR = 0.734) and Katz (AUROC = 0.724 and AUPR = 0.741) (Fig. 2a and b, Supplementary Table S5).

Since the number of correctly predicted true positives reflects the discriminatory power of a prediction method to distinguish true positives, especially when the number of negative samples is far larger than that of positive samples, we further used ‘recall @ top- $k$ ’ as the

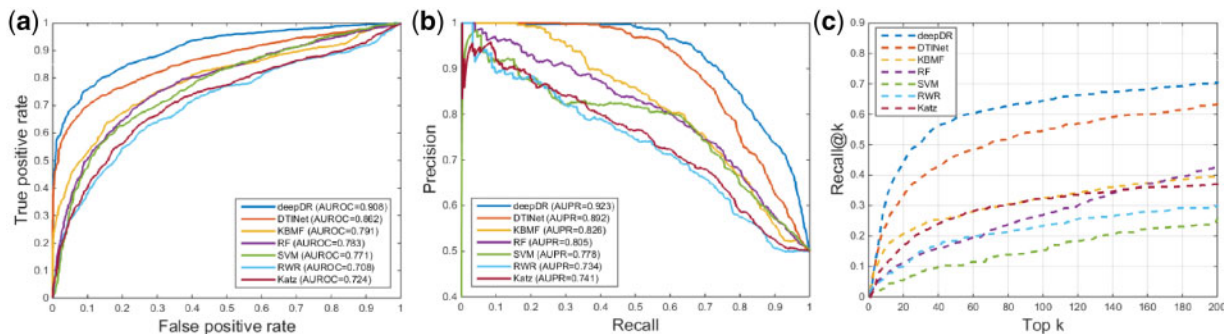
evaluation metric, which is defined as the fraction of true approved diseases (indication) that were retrieved in the list of top- $k$  predictions for a drug. The motivation of using this metric is that a method that can accurately recover the true interacting diseases in the list of top- $k$  predictions is generally desired and useful for downstream experimental validation. As shown in Figure 2c, recall of deepDR was 70% for top 200 predicted candidates, significantly outperforming that of DTINet (62%), KBMF (40%), RF (43%), SVM (25%), RWR (30%) and Katz (38%).

### 3.4 Performance of deepDR on the external validation set

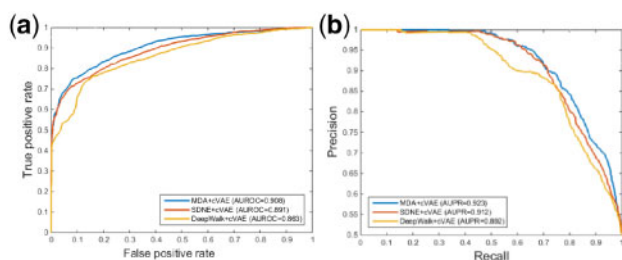
Cross-validation on retrospective data probably leads to overoptimistic results. For objective performance evaluation, we further collected clinically reported drug–disease pairs from the ClinicalTrials.gov database, as an external validation set. The external validation set carries 129 newly discovered associations that were not used before. Figure 3a gives the performance comparison of all the methods for this task. deepDR achieved superior performance over the other methods. For example, deepDR achieved AUROC value of 0.826, significantly outperforming that of DTINet (0.732), KBMF (0.673), RF (0.788), SVM (0.753), RWR (0.601) and Katz (0.563) as well. Moreover, we further showed the recall rate of correctly predicted drug–disease associations with respect to given top-ranked thresholds, as shown in Figure 3b. Specifically, deepDR (recall = 65%) achieved more correctly predicted drug–disease associations than other methods on almost every top-rank 200 threshold, consistent with the 5-fold cross-validation, indicating a high generalization ability.



**Fig. 3.** Evaluation of deepDR on the external validation set collected from the ClinicalTrials.gov database (see Section 2). (a) Receiver operating characteristic (ROC) curves of prediction results obtained by applying deepDR and other competitive methods. (b) Recall against top  $k$  predicted list in the external validation



**Fig. 2.** Performance of different methods on the clinically reported drug–disease network. (a) Receiver operating characteristic (ROC) curves of prediction results obtained by applying deepDR and six previously reported methods in 5-fold cross-validation. (b) Precision–recall (PR) curves of prediction results obtained by applying deepDR and other competitive methods in 5-fold cross-validation. (c) Recall of deepDR and other methods against top  $k$  predicted list during 5-fold cross-validation



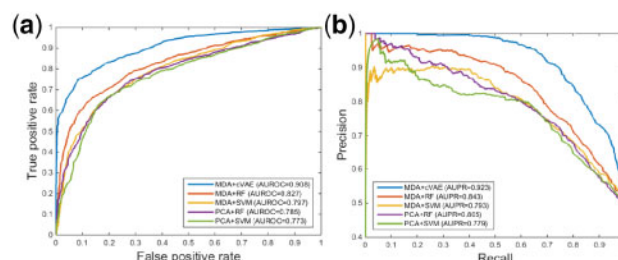
**Fig. 4.** Performance of MDA when comparing with different network representation approaches. (a) Receiver operating characteristic (ROC) curves of prediction results obtained by applying deepDR and other methods. (b) Precision–recall (PR) curves of prediction results obtained by applying deepDR and other methods

### 3.5 Performance of deepDR by ablation analysis

deepDR mainly consists of two parts, namely MDA and cVAE. To examine the contribution of each component, we compared deepDR with several combinations. First, we compared MDA with two other network representation approaches, DeepWalk (Perozzi *et al.*, 2014) and SDNE (Wang *et al.*, 2016), to inspect the contribution of MDA. DeepWalk transforms a graph into collections of linear sequences by truncated random walk, and utilizes the skip-gram model to learn low-dimensional representations for vertices (Perozzi *et al.*, 2014). SDNE is a deep model for network representation, which considers the first and the second order proximities using a deep auto-encoder (Wang *et al.*, 2016). As shown in Figure 4a and b, we found that MDA outperformed that of DeepWalk and SDNE (Fig. 4a and b). Specifically, both MDA and SDNE are deep learning-based embedding methods and they both outperformed DeepWalk. As published previously, SDNE cannot consider the inner interaction of different networks. In this study, MDA was able to fuse different information into a low-dimensional feature representation common to all networks and achieved the best performance in this experiment. To inspect the contribution of cVAE further, we compared deepDR with SVM and RF using the same features extracted from MDA or traditional principal components analysis (PCA). Specifically, we compared deepDR with four different combinations, including MDA + RF, MDA + SVM, PCA + RF, PCA + SVM, the evaluation results of these combinations are reported in Figure 5a and b. We found that deepDR achieved the best performance. Owing to lacking of disease-related information, traditional classifiers did not exhibit excellent performance, while deepDR made full use of multiple drug-related information and did not need disease-related networks, which accounts for its superior performance over other methods.

### 3.6 Pharmacological interpretation of deepDR

Deep neural network-based network embedding is able to encode the complex feature relationships relevant for the predictive task into the vector representations of the vertex. One intuitionistic way of assessing the quality of the vertex representations is through visualization, much of knowledge embedded within the learned features is encoded by the feature matrix. By visualizing the feature matrix, we can disentangle the complex network. We examined the internal features learned by network embedding using *t*-SNE (*t*-distributed stochastic neighbor embedding) (van der Maaten and Hinton, 2008), a non-linear dimensionality reduction method that embeds similar points in the high-dimensional space as points close in two dimensions. Each point in the figure represents a drug node projected from the 900-dimensional feature vectors extracted from the



**Fig. 5.** Performance of cVAE when comparing with different traditional classifiers. (a) Receiver operating characteristic (ROC) curves of prediction results obtained by comparing deepDR with other methods. (b) Precision–recall (PR) curves of prediction results obtained by comparing deepDR with other methods

middle layer of MDA in the deepDR framework. The same type of nodes is highlighted with the same color. Under the same setting, a clustering with clear boundaries between different color groups indicate better representations. Via *t*-SNE, we projected drugs grouped by the first-level of the anatomical therapeutic chemical (ATC) classification system code into 2D space. We found that the feature vectors generated by MDA are able to distinguish 14 types of drugs grouped by ATC codes very well (Supplementary Fig. S1a), significantly outperforming PCA approach (Supplementary Fig. S1b). Taken together, deepDR indicates high network embedding capability to preserve the inherent properties and structures of the heterogeneous drug–disease network.

### 3.7 Case study: computationally identified approved drugs for Alzheimer’s disease and Parkinson’s disease

To further validate the prediction ability of deepDR, we conducted a case study for two types of neurodegenerative diseases which do not have efficacious treatments available yet, including Alzheimer’s disease (AD) and Parkinson’s disease (PD).

**Alzheimer’s disease (AD).** We focused on the top 20 deepDR-predicted candidates for AD in Supplementary Table S3. For each drug, we showed the canonical name, predicted score and the literature-reported evidence. Isoprenaline, a non-selective  $\beta$ -adrenoreceptor agonist, was approved for the treatment of bradycardia and heart block (Sato *et al.*, 2004). Herein, isoprenaline is the top first predicted candidate for potentially treating AD. A previous clinical studies reported that isoprenaline may reduce amyloid plaques in Alzheimer’s patients (Ohm *et al.*, 1991). Dopamine, a compound of the catecholamine and phenethylamine families playing important roles in the human brain, was predicted by deepDR to be associated with AD. Such a prediction can be supported by a previous study indicating that lack of dopamine in the brain may cause some of the earliest symptoms of Alzheimer (Li *et al.*, 2004). Risperidone, an atypical antipsychotic, which was primarily used in the treatment of schizophrenia and bipolar disorder, was predicted by deepDR to also have potential effect on Alzheimer’s disease. This prediction is supported by previous literatures (Katz *et al.*, 2007; Negron and Reichman, 2000). In addition, deepDR found that aripiprazole, another atypical antipsychotic primarily used in the treatment of schizophrenia and bipolar disorder, was associated with AD, which was supported by several evidences (De Deyn *et al.*, 2005, 2013). Among the top 20 predicted drugs ranked according to their confidence scores, 14 drugs (70% success rate) were validated by various evidences from clinical studies, preclinical studies and other literature data (Supplementary Table S3).

**Parkinson's disease (PD).** We focused on the top 20 deepDR-predicted candidates on PD in [Supplementary Table S4](#). We found that 14 of 20 drugs (70% success rate) were validated by previous studies from literatures ([Supplementary Table S4](#)). For instance, orphenadrine, an anticholinergic drug of the ethanolamine antihistamine class, was predicted by deepDR to have an association with PD. Such a prediction can be supported by previous literatures ([Bassi et al., 1986](#); [Strang, 1964](#)). Pergolide, an ergoline-based dopamine receptor agonist associated with reduced dopamine activity in the substantia nigra of the brain, was predicted by deepDR to be associated with PD, which is consistent with previous preclinical and clinical studies ([Goetz et al., 1983](#); [Storch et al., 2005](#); [Van Camp et al., 2004](#)). In addition, methylphenidate, a stimulant medication used to treat attention deficit hyperactivity disorder (ADHD) and narcolepsy, was predicted to associate with PD, supported by multiple studies ([Auriel et al., 2009](#); [Devos et al., 2013](#); [Mendonça et al., 2007](#)).

In summary, deepDR offers a useful tool to prioritize potential repurposed drugs for Alzheimer's disease and Parkinson's disease.

## 4 Discussion and conclusions

In this study, we presented a novel deep learning framework deepDR to uncover the potential associations between drugs and diseases. Apart from the gold standard drug–disease association network, we integrated one drug–drug interaction network, one drug–protein association network, one drug–side-effect association network and six drug–drug similarity networks to construct a complicated heterogeneous network which contains diverse information and a multi-perspective view for predicting novel drug–disease associations. deepDR first fuses diverse information from a multitude of different network types into a compact, low-dimensional feature representation and then the learned low-dimensional representation of drug features together with known drug–disease interaction pairs are fed into a variational autoencoder to predict new drug–disease associations. Theoretically, deepDR is superior to the existing drug repositioning methods as we adopt a multi-modal deep autoencoder (MDA) to capture complex topological patterns across different data sources. Further, deepDR is able to preserve the non-linear network structure by applying multiple layers of non-linear functions. It also complements the sparse ratings with drug features, as feeding side information into the same VAE increases the number of samples for training, which also acts like a pre-training step. Specifically, drug–disease associations and drug features are different sources of information, both are information with drugs, so they can be encoded and decoded collectively through the same inference network and generation network. We have validated the prediction ability of deepDR in terms of cross validation, external validation and case studies, and the results show that our method achieves state-of-the-art performance for the discovery of new drug–disease associations. The deepDR makes full use of multiple drug-related networks, whereas other methods like RWR and Katz all need one drug similarity network and one disease similarity network. To compare our method with the other methods, we calculated a disease similarity network from the Jaccard similarity of disease–protein association network, and chose drug chemical similarity for these methods, so there may be some deviation during the comparison. In future studies, since deepDR is a scalable framework, collecting and incorporating more relevant association data from more databases and the literature may improve its power. In addition, owing to the structure of cVAE, at present deepDR can only integrate drug-

related information, a future direction is to modify the structure of neural network recommendation and extend deepDR to integrate both drug-related information and disease-related information.

We acknowledge several potential limitations of deepDR under current network-based deep learning framework. Although our sizeable efforts assembled large-scale, experimentally reported drug–target interactions from publicly available databases, data quality is also not assured and the network data may be incomplete. For example, owing to the lack of negative drug–disease pairs in the publicly available databases and published literatures, it has been challenging to build gold-standard unknown pairs as negative samples during machine learning studies. We provided the entire lists of top deepDR-predicted drug–disease pairs in [Supplementary Database S1](#). State-of-the-art pharmaco-epidemiologic analysis on patient data (e.g. health insurance claims data) and *in vitro* or *in vivo* mechanistic studies for the deepDR-predicted candidates are warranted in the future.

In summary, our findings suggest that *in silico* drug repurposing could benefit from network-based, deep learning, exploring the relationships of drug–target–disease heterogeneous networks. From a translational perspective, the network tools developed here could help develop novel, efficacious therapies from network-based drug repurposing perspectives for multiple complex diseases if broadly applied.

## Funding

This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health [K99HL138272 and R00HL138272 to F.C.]. This work has been also supported in part with Federal funds from the Frederick National Laboratory for Cancer Research, National Institutes of Health [HHSN261200800001E]. This research was supported (in part) by the Intramural Research Program of NIH, Frederick National Lab, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

*Conflict of Interest:* none declared.

## References

- Angermueller, C. et al. (2016) Deep learning for computational biology. *Mol. Syst. Biol.*, **12**, 878.
- Auriel, E. et al. (2009) Methylphenidate for the treatment of Parkinson disease and other neurological disorders. *Clin. Neuropharmacol.*, **32**, 75–81.
- Avorn, J. (2015) The \$2.6 billion pill—methodologic and policy considerations. *N. Engl. J. Med.*, **372**, 1877–1879.
- Bassi, S. et al. (1986) Treatment of Parkinson's disease with orphenadrine alone and in combination with L-dopa. *Br. J. Clin. Pract.*, **40**, 273–275.
- Bodenreider, O. (2004) The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res.*, **32**, D267–270.
- Breiman, L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.
- Brown, A.S. and Patel, C.J. (2017) A standard database for drug repositioning. *Sci. Data*, **4**, 170029.
- Bullinaria, J.A. and Levy, J. (2007) Extracting semantic representations from word co-occurrence statistics: a computational study. *Behav. Res. Methods*, **39**, 510–526.
- Cao, M. et al. (2014) New directions for diffusion-based network prediction of protein function: incorporating pathways with confidence. *Bioinformatics*, **30**, i219–227.
- Cao, S. et al. (2016) Deep neural network for learning graph representations. In: *Thirteenth AAAI Conference on Artificial Intelligence*, AAAI Publications, pp. 1145–1152. Phoenix, AZ, USA.

- Chen, Y. and De Rijke, M. (2018) A collective variational autoencoder for top-N recommendation with side information. arXiv: 1807.05730.
- Cheng, F. et al. (2018) Network-based approach to prediction and population-based validation of *in silico* drug repurposing. *Nat. Commun.*, **9**, 2691.
- Cheng, F. et al. (2017) Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief. Bioinformatics*, **18**, 682–697.
- Cheng, F. et al. (2019) Network-based prediction of drug combinations. *Nat. Commun.*, **10**, 1197.
- Cheng, F. et al. (2012) Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.*, **8**, e1002503.
- Ching, T. et al. (2018) Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface*, **15**, 20170387.
- Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Mach. Learn.*, **20**, 273–297.
- De Deyn, P. et al. (2005) Aripiprazole for the treatment of psychosis in patients with Alzheimer's disease: a randomized, placebo-controlled study. *J. Clin. Psychopharmacol.*, **25**, 463–467.
- De Deyn, P. et al. (2013) Aripiprazole in the treatment of Alzheimer's disease. *Exp. Opin. Pharmacother.*, **14**, 459–474.
- Devos, D. et al. (2013) Methylphenidate: a treatment for Parkinson's disease? *CNS Drugs*, **27**, 1–14.
- Elkan, C. and Noto, K. (2008) Learning classifiers from only positive and unlabeled data. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2008, pp. 213–220. Las Vegas, NV, USA.
- Gligorijevic, V. et al. (2018) deepNF: deep network fusion for protein function prediction. *Bioinformatics*, **34**, 3873–3881.
- Goetz, C.G. et al. (1983) Pergolide in Parkinson's disease. *Arch. Neurol.*, **40**, 785–787.
- Gönen, M. et al. (2013) Kernelized Bayesian matrix factorization. Preprint at: arXiv, 1211.1275.
- Greene, J.A. and Loscalzo, J. (2017) Putting the patient back together—social medicine, network medicine, and the limits of reductionism. *N. Engl. J. Med.*, **377**, 2493–2499.
- Hsieh, C.J. et al. (2014) PU learning for matrix completion. Preprint at: arXiv, 1411.6081.
- Katz, I. et al. (2007) The efficacy and safety of risperidone in the treatment of psychosis of Alzheimer's disease and mixed dementia: a meta-analysis of 4 placebo-controlled clinical trials. *Int. J. Geriatr. Psychiatry*, **22**, 475–484.
- Kingma, D.P. and Welling, M. (2013) Auto-encoding variational Bayes. Preprint at: arXiv, 1312.6114.
- Köhler, S. et al. (2008) Walking the interactome for prioritization of candidate disease genes. *Am. J. Hum. Genet.*, **82**, 949–958.
- Law, V. et al. (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.*, **42**, D1091–D1097.
- Li, J.I.E. et al. (2004) Dopamine and L-dopa disaggregate amyloid fibrils: implications for Parkinson's and Alzheimer's disease. *FASEB J.*, **18**, 962–964.
- Luo, Y. et al. (2017) A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.*, **8**, 573.
- Mendonça, D.A. et al. (2007) Methylphenidate improves fatigue scores in Parkinson disease: a randomized controlled trial. *Mov. Disord.*, **22**, 2070–2076.
- Nagarajan, N. and Dhillon, I.S. (2014) Inductive matrix completion for predicting gene-disease associations. *Bioinformatics*, **30**, i60–i68.
- Negron, A.E. and Reichman, W.E. (2000) Risperidone in the treatment of patients with Alzheimer's disease with negative symptoms. *Int. Psychogeriatr.*, **12**, 527–536.
- Ohm, T.G. et al. (1991) Reduced basal and stimulated (isoprenaline, Gpp(NH)p, forskolin) adenylate cyclase activity in Alzheimer's disease correlated with histopathological changes. *Brain Res.*, **540**, 229–236.
- Paisley, J. et al. (2012) Variational Bayesian inference with stochastic search. Preprint at: arXiv, 1206.6430.
- Perozzi, B. et al. (2014) DeepWalk: online learning of social representations. In: *KDD' 14 Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 701–710, Preprint at: arXiv: 1403.6652.
- Pushpakom, S. et al. (2018) Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.*, **18**, 41.
- Sato, M. et al. (2004) Loss of beta-adrenoceptor response in myocytes overexpressing the Na<sup>+</sup>/Ca<sup>2+</sup>-exchanger. *J. Mol. Cell Cardiol.*, **36**, 43–48.
- Storch, A. et al. (2005) High-dose treatment with pergolide in Parkinson's disease patients with motor fluctuations and dyskinesias. *Parkinsonism Relat. Disord.*, **11**, 393–398.
- Strang, R.R. (1964) Orphenadrine in the treatment of Parkinson's disease. *Curr. Med. Drugs*, **5**, 24–31.
- Tan, S.Y. and Grimes, S. (2010) Paul Ehrlich (1854–1915): man with the magic bullet. *Singapore Med. J.*, **51**, 842–843.
- Topol, E.J. (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.*, **25**, 44–56.
- Singh-Blom, U.M. et al. (2013) Prediction and validation of gene-disease associations using methods inspired by social network analyses. *PLoS One*, **8**, e58977.
- Van Camp, G. et al. (2004) Treatment of Parkinson's disease with pergolide and relation to restrictive valvular heart disease. *Lancet*, **363**, 1179–1183.
- van der Maaten, L. and Hinton, G. (2008) Visualizing data using t-SNE. *J. Mach. Learn. Res.*, **9**, 2579–2605.
- Vincent, P. et al. (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, **11**, 3371–3408.
- Wang, D. et al. (2016) Structural deep network embedding. In: *KDD' 16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1225–1234. San Francisco, CA, USA.
- Wishart, D.S. et al. (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
- Xing, E.P. et al. (2003) A generalized mean field algorithm for variational inference in exponential families. In: *UAI'03 Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, pp. 583–591. Preprint at: arXiv: 1212.2512.