Li *et al. BMC Bioinformatics* (2024) 25:261
https://doi.org/10.1186/s12859-024-05893-5

BMC Bioinformatics

RESEARCH Open Access

# Drug repositioning based on residual attention network and free multiscale adversarial training

Guanghui Li[1*], Shuwen Li[1], Cheng Liang[2], Qiu Xiao[3] and Jiawei Luo[4*]

*Correspondence:
ghli16@hnu.edu.cn;
luojiawei@hnu.edu.cn

[1] School of Information Engineering, East China Jiaotong University, Nanchang, China
[2] School of Information Science and Engineering, Shandong Normal University, Jinan, China
[3] College of Information Science and Engineering, Hunan Normal University, Changsha, China
[4] College of Computer Science and Electronic Engineering, Hunan University, Changsha, China

## Abstract

**Background:** Conducting traditional wet experiments to guide drug development is an expensive, time-consuming and risky process. Analyzing drug function and repositioning plays a key role in identifying new therapeutic potential of approved drugs and discovering therapeutic approaches for untreated diseases. Exploring drug-disease associations has far-reaching implications for identifying disease pathogenesis and treatment. However, reliable detection of drug-disease relationships via traditional methods is costly and slow. Therefore, investigations into computational methods for predicting drug-disease associations are currently needed.

**Results:** This paper presents a novel drug-disease association prediction method, RAFGAE. First, RAFGAE integrates known associations between diseases and drugs into a bipartite network. Second, RAFGAE designs the Re_GAT framework, which includes multilayer graph attention networks (GATs) and two residual networks. The multilayer GATs are utilized for learning the node embeddings, which is achieved by aggregating information from multihop neighbors. The two residual networks are used to alleviate the deep network oversmoothing problem, and an attention mechanism is introduced to combine the node embeddings from different attention layers. Third, two graph autoencoders (GAEs) with collaborative training are constructed to simulate label propagation to predict potential associations. On this basis, free multiscale adversarial training (FMAT) is introduced. FMAT enhances node feature quality through small gradient adversarial perturbation iterations, improving the prediction performance. Finally, tenfold cross-validations on two benchmark datasets show that RAFGAE outperforms current methods. In addition, case studies have confirmed that RAFGAE can detect novel drug-disease associations.

**Conclusions:** The comprehensive experimental results validate the utility and accuracy of RAFGAE. We believe that this method may serve as an excellent predictor for identifying unobserved disease-drug associations.

**Keywords:** Graph attention network, Residual network, Graph autoencoder, Adversarial training, Drug-disease association

## Background

Drugs play important roles in treating diseases and promoting the health of organisms [1]. However, traditional drug development is an extremely lengthy and expensive process [2]. Recent studies have estimated that the average development cost to approve a new drug is $2.6 billion and the average development time is 10 years [3]. Drug repositioning, which involves discovering new therapeutic outcomes for previously approved drugs, is considered an important alternative to traditional drug development [4–8]. This approach shortens drug development and research cycles to 7 years, reduces costs to $295 million, and is more reliable than novel drug development [9]. Therefore, using known drugs for new disease treatments is gaining popularity [10, 11]. Traditional methods of discovering abnormal clinical manifestations through manual screening of clinical drug databases requires extensive experimentation. With the continuous accumulation of a wide variety of biological data, numerous computational methods based on data mining techniques have gained traction [12].

Matrix factorization aims to approximate the initial matrix by decomposing it into the product of two low-rank matrices, which are represented by hidden factor vectors in the $k$-dimension. The inner product of the drug and disease vectors represents the association between them. Previous studies have shown that matrix decomposition methods are effective computational methods for drug-disease association prediction [13–17]. For example, the similarity constrained matrix factorization method for the drug-disease association prediction (SCMFDD) method, proposed by Zhang et al., maps the associations between diseases and drugs into two low-ranking spaces and reveals the basic features. Then, drug similarity and disease similarity are introduced as increasing constraints [18]. Furthermore, Yang et al. proposed the multisimilarities bilinear matrix factorization (MSBMF) approach, which connects multiple disease and drug similarity matrices and extracts the effective latent features in the similarity matrix to infer associations between diseases and drugs [19]. In addition, Zhang et al. proposed a new drug repositioning method by using Bayesian inductive matrix completion (DRIMC), which uses the complement of Bayesian inductive matrices. This method integrates multiple similarities into a fused similarity matrix, where similarity information is described by similarity values between a drug or disease and its $k$-nearest neighbors. Finally, the disease-drug association is predicted via induction matrix completion [20].

Networks can represent the complex relationships among entities, and the methods used to construct biological networks can effectively utilize information from multiple biological entities to represent the degree of association between them [21]. The network-based method has produced good results in drug repositioning [22–24]. For instance, Zhao et al. first constructed a heterogeneous information network by combining drug-disease, protein-disease and drug-protein bioinformatics networks with disease and drug biology information. Then, the combined features of the nodes were learned from a biological and topological perspective via different representations. Moreover, random forest classifiers can be used to predict unknown associations [25]. Zhang et al. proposed a multiscale neighborhood topology learning method for drug repositioning (MTRD) to learn and integrate multiscale neighborhood topologies. This method involves the construction of different drug-disease heterogeneous networks to discover new drug-disease associations [26]. In addition, Luo et al. proposed

a method named MBiRW that uses similarity matrices and known associations to construct heterogeneous networks and predicts unknown associations via the double random walk algorithm [27].

Although matrix factorization methods achieve good performance, they are weak in the interpretability of associations between diseases and drugs, whereas network methods are biased in representing higher-order networks. To solve these problems, several pioneering studies have focused on developing deep learning-based drug repositioning models [28–33]. For example, Zeng et al. first integrated multiple disease-drug biological networks and designed a multimodal deep autoencoder named deep learning-based drug repositioning (deepDR) for learning higher order neighborhood information of drug-disease associations [34]. Subsequently, Yu et al. constructed a graph convolutional network (GCN) architecture with attention mechanisms, i.e., the label-aware GCN (LAGCN). First, this method uses known drug-disease associations, disease-disease similarities and drug-drug similarities to construct heterogeneous networks and applies GCNs to the network. Next, the embeddings from multiple GCN layers are integrated via layer attention mechanisms. Finally, drug-disease pairs are scored on the basis of the integrated embeddings [35]. Feng et al. proposed Protein And Drug Molecule interaction prEdiction (PADME), a novel method to combine molecular GCNs for compound featurization with protein descriptors for drug-target interaction prediction [36]. Moreover, Meng et al. proposed a drug repositioning approach based on weighted bilinear neural collaborative filtering (DRWB-NCF) on the basis of neighborhood interaction and collaborative filtering. Instead of using all neighbors, this method uses only the nearest neighbors, thus filtering out noise and yielding more precise results [37]. Recently, Gu et al. proposed a method named relations-enhanced drug-disease association (REDDA) for learning node features of heterogeneous networks and topological subnetworks. This method employs heterogeneous networks as the backbone and combines the backbone with three attention mechanisms [38]. Deep learning-based methods mainly construct heterogeneous networks by using supplementary information about diseases and drugs and learn the features of diseases and drugs by applying deep learning algorithms to these networks.

However, these deep learning-based approaches tend to have oversmoothing problems caused by the homogenization of node embeddings and are highly dependent on the input quality. In this paper, we present a novel method of drug repositioning named RAFGAE. This method combines residual networks, graph attention networks (GATs), graph autoencoders (GAEs) and adversarial training to predict unknown associations between diseases and drugs. First, we use disease semantic similarity, drug structural similarity and disease-drug associations to construct the initial input features. GATs are used to facilitate the learning of disease and drug embeddings in each layer and combine the embedding of different layers via attention mechanisms. Moreover, the initial residual and adaptive residual connections are adopted to alleviate the oversmoothing problem. Then, two GAEs are constructed on the basis of the disease space and drug space, and the information in these spaces can be integrated through synergistic training. Finally, the scores of the two GAEs are linearly combined by a balancing parameter to calculate the final prediction scores. On this basis, adversarial training is introduced

to reduce invalid information and data noise, improving the input quality. The main contributions of RAFGAE can be summarized as follows:

- RAFGAE is a complete deep learning approach that can effectively predict the associations between diseases and drugs.
- RAFGAE designs the Re_GAT framework, which includes multilayer GATs and two residual networks. Multilayer GATs are utilized to learn the node embeddings by aggregating information from multihop neighbors, and two residual networks are used to alleviate the deep network oversmoothing problem. Then, an attention mechanism is introduced to combine the node embeddings of different attention layers.
- RAFGAE performs adversarial training that may eliminate abnormal values, missing values and noise, increasing the input quality and prediction accuracy when extracting associations between diseases and drugs.
- Our comprehensive experimental results demonstrate that the proposed RAFGAE method significantly outperforms five state-of-the-art methods on the benchmark dataset.

## Results and discussion

### Algorithm performance comparison

To verify the performance of RAFGAE, we compare it with five recently proposed methods.

- DRWBNCF [37], a method for drug repositioning on the basis of neighborhood interaction and collaborative filtering, uses only the nearest neighbors, rather than all neighbors, to filter out noisy information. A new weighted bilinear GCN encoder is then proposed.
- LAGCN [35], a layer attention GCN method for drug repositioning, encodes a heterogeneous network combining known drug-disease associations, disease similarity and drug similarity information. To integrate all useful information, a layer attention mechanism is introduced into multiple GCN layers.
- In bounded nuclear norm regularization (BNNR) [39], a heterogeneous network is constructed. This network combines known drug-disease associations, disease similarity and drug similarity information. The method tolerates noise by adding a regularization term to balance the rank properties and approximation error.
- The neural inductive matrix completion with GCN (NIMCGCN) method [40], a method for the prediction of miRNA-disease associations) first employs GCN to learn the features of diseases and miRNAs from the disease and miRNA similarity networks. Then, neural induction matrix completion is applied for association matrix completion.
- SCMFDD [18] (a similarity constraint matrix completion method for the prediction of drug-disease associations) projects known drug-disease association information into two low-rank spaces, revealing potential disease and drug embeddings, and then introduces drug featured-based and disease semantic similarities as constraints for drugs and diseases in the low-rank spaces.

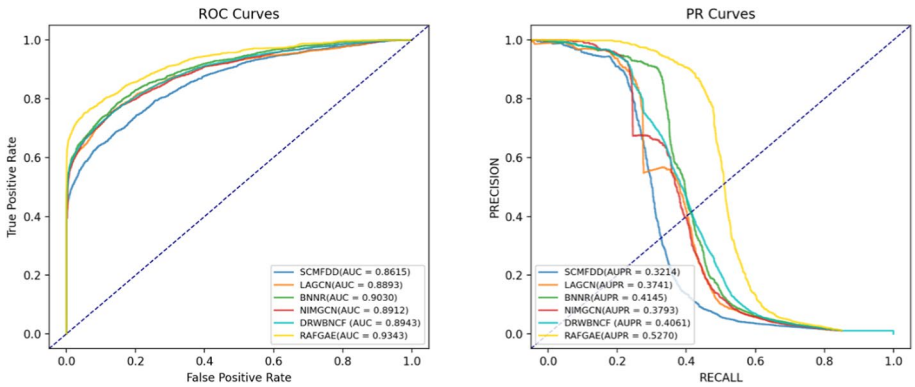


**Fig. 1** ROC curves and PR curves of RAFGAE and other models on the F-dataset

**Table 1** Performance of the comparison methods on the two datasets

| Dataset | Method | AUC | AUPR | F1-score | MCC |
|---|---|---|---|---|---|
| F-dataset | DRWBNCF | 0.8943 ± 0.0020 | 0.4061 ± 0.0064 | 0.4702 ± 0.0040 | 0.5222 ± 0.0016 |
| | LAGCN | 0.8893 ± 0.0026 | 0.3741 ± 0.0048 | 0.4003 ± 0.0036 | 0.5145 ± 0.0040 |
| | BNNR | 0.9030 ± 0.0046 | 0.4145 ± 0.0036 | 0.4467 ± 0.0034 | 0.5478 ± 0.0044 |
| | NIMCGCN | 0.8912 ± 0.0042 | 0.3793 ± 0.0036 | 0.4019 ± 0.0016 | 0.4640 ± 0.0054 |
| | SCMFDD | 0.8615 ± 0.0026 | 0.3214 ± 0.0034 | 0.4080 ± 0.0020 | 0.5095 ± 0.0026 |
| | RAFGAE | 0.9343 ± 0.0060 | 0.5270 ± 0.0314 | 0.5557 ± 0.0060 | 0.6084 ± 0.0050 |
| C-dataset | DRWBNCF | 0.8951 ± 0.0026 | 0.4047 ± 0.0042 | 0.4722 ± 0.0062 | 0.5230 ± 0.0048 |
| | LAGCN | 0.8901 ± 0.0040 | 0.3692 ± 0.0034 | 0.4019 ± 0.0042 | 0.5159 ± 0.0042 |
| | BNNR | 0.9052 ± 0.0052 | 0.4140 ± 0.0042 | 0.4504 ± 0.0040 | 0.5502 ± 0.0040 |
| | NIMCGCN | 0.8924 ± 0.0026 | 0.3789 ± 0.0036 | 0.4103 ± 0.0046 | 0.4683 ± 0.0026 |
| | SCMFDD | 0.8637 ± 0.0020 | 0.3216 ± 0.0044 | 0.4124 ± 0.0048 | 0.5124 ± 0.0034 |
| | RAFGAE | 0.9346 ± 0.0060 | 0.5237 ± 0.0316 | 0.5670 ± 0.0050 | 0.6223 ± 0.0026 |

The above methods also involve similarity-based graph neural network models. The parameters in these methods are set to either the optimal values via a grid search (for DRWBNCF, $\lambda$ is selected from {0.1, 0.2, ..., 0.9}; for BNNR, $\alpha$ and $\beta$ are chosen from {0.01, 0.1, 1, 10}; and for SCMFDD, $k$ is selected from{5%, 10%, ..., 50%}) or the values recommended by the authors (for LAGCN, $\alpha = 4000$, $\beta$ =0.6, and $\gamma = 0.4$; and for NIMCGCN, $\alpha = 0.4$, $l = 3$, and $t = 2$). Furthermore, to ensure a meaningful and relevant comparison, each of the comparison methods is initially evaluated via the same 10-fold cross-validation approach and on the same benchmarking sets as those for our proposed method, RAFGAE. This approach allows us to conduct a comprehensive and rigorous assessment of the performance of all the methods.

The area under the curve (AUC) values in Fig. 1 and Table 1 show a comparison of the model performance. On the F-dataset, RAFGAE achieves the highest AUC score of 0.9343, which is 7.28%, 4.50%, 3.13%, 4.31%, and 4.01% higher than those of SCMFDD, LAGCN, BNNR, NIMGCN, and DRWBNCF, respectively. Similarly, on the C-dataset, RAFGAE achieves the highest AUC score of 0.9346. By comparing the model proposed in this paper with other models, it is evident that introducing residual connections and adversarial training can enhance the predictive performance of our model. Overall,

the above experiments show that RAFGAE is an excellent predictor of disease-drug relationships.

### Ablation study

To quantitatively evaluate the importance of the two modules (the Re_GAT framework and the FMAT module) to RAFGAE, ablation experiments are conducted. The details of these variants of RAFGAE are listed below:

- RAFGAE: The comprehensive RAFGAE framework consists of three main components: the Re_GAT framework, the FMAT module, and the GAE module.
- GAE: The RAFGAE variant that includes only the GAE module.
- FGAE: The RAFGAE variant that includes the FMAT and GAE modules but excludes the Re_GAT framework.
- RAGAE: The RAFGAE variant that includes Re_GAT framework and the GAE module but excludes the FMAT module.

According to Fig. 2 and Table 2, it is clear that RAFGAE achieved the highest AUC and area under the precision–recall (AUPR) curve values on both the F-dataset and the C-dataset. The RAGAE and FGAE results show the impacts of global neighborhood node information aggregation and adversarial feature enhancement on the RAFGAE performance, respectively. In addition, the GAE results demonstrate that combining the Re_GAT framework and the FMAT module can improve the predictive performance of the RAFGAE model. In comparing FGAE and RAGAE to GAE, the performance results imply that both the Re_GAT framework and the FMAT module can improve the model performance. The poor performance of GAE suggests that the use of multilayer attention networks to aggregate global information and the incorporation of residual architectures to address the potential oversmoothing problem can enhance the accuracy of drug-disease association prediction. Furthermore, the results indicate that the inclusion of the

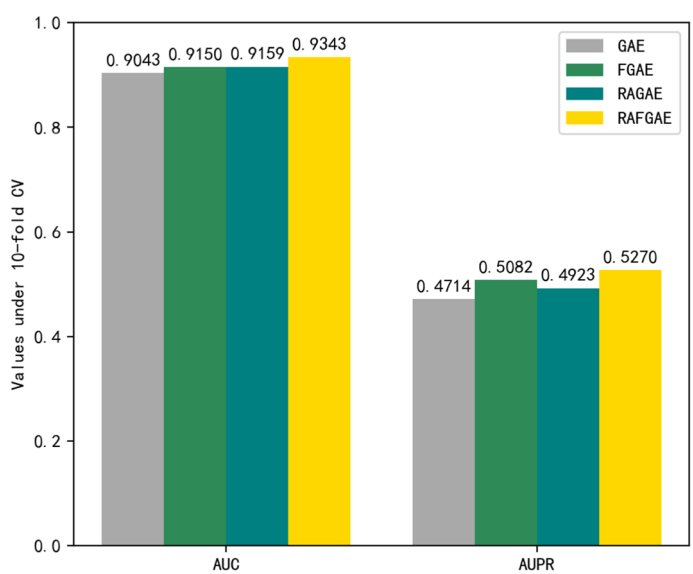


**Fig. 2** Results of RAFGAE and its variants in the ablation study on the F-dataset

**Table 2** Performance comparison between RAFGAE and its variants

| Dataset | Method | AUC | AUPR | F1-score | MCC |
|---|---|---|---|---|---|
| F-dataset | GAE | 0.9043 ± 0.0040 | 0.4714 ± 0.0040 | 0.5021 ± 0.0024 | 0.5322 ± 0.0016 |
| | FGAE | 0.9150 ± 0.0010 | 0.5082 ± 0.0026 | 0.5303 ± 0.0026 | 0.5715 ± 0.0022 |
| | RAGAE | 0.9159 ± 0.0070 | 0.4923 ± 0.0034 | 0.5167 ± 0.0042 | 0.5578 ± 0.0024 |
| | RAFGAE | 0.9343 ± 0.0060 | 0.5270 ± 0.0314 | 0.5557 ± 0.0060 | 0.6084 ± 0.0050 |
| C-dataset | GAE | 0.9049 ± 0.0046 | 0.4728 ± 0.0036 | 0.5225 ± 0.0060 | 0.5343 ± 0.0028 |
| | FGAE | 0.9212 ± 0.0034 | 0.5123 ± 0.0026 | 0.5427 ± 0.0042 | 0.5720 ± 0.0036 |
| | RAGAE | 0.9155 ± 0.0042 | 0.4980 ± 0.0040 | 0.5301 ± 0.0016 | 0.5592 ± 0.0024 |
| | RAFGAE | 0.9346 ± 0.0060 | 0.5309 ± 0.0316 | 0.5670 ± 0.0050 | 0.6223 ± 0.0026 |

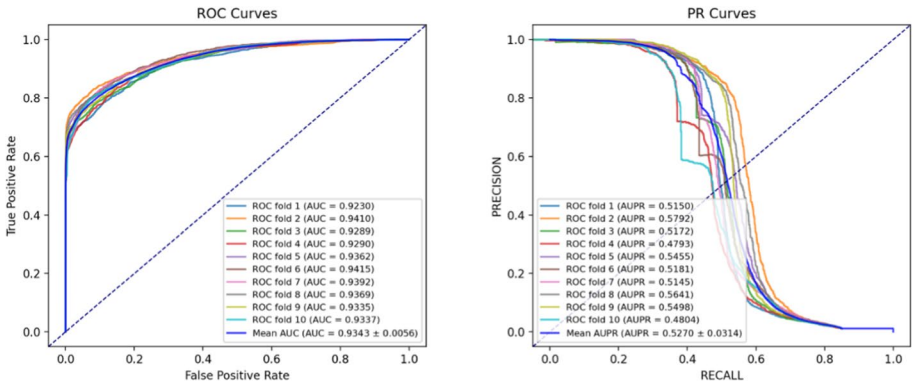


**Fig. 3** RAFGAE ROC and PR curves via tenfold CV on the F-dataset

adversarial training module improves the input quality, thereby satisfying the requirements of deep neural networks for high-quality input features. These results demonstrate that the RAFGAE structure is reasonable.

### Performance evaluation

To assess the effectiveness of RAFGAE in predicting known associations, tenfold cross validation (CV) is applied. In tenfold CV, the dataset is divided into ten folds. Nine folds are used as the training set, and the remaining fold is used to validate the performance of RAFGAE. This process is repeated 10 times, with each fold used as the testing fold once. Several important indicators are used to evaluate the performance of RAFGAE. The receiver operating characteristic (ROC) curve, which is based on the false-positive rate (FPR) and the true positive rate (TPR), is utilized. As the benchmark datasets used in this experiment are imbalanced, we also use the PR curve and calculate the area under the PR curve (AUPR) as two additional indicators. To further evaluate the overall performance of the prediction model from multiple perspectives, the F1 score and the Mathews correlation coefficient (MCC) are calculated.

The ROC and PR curves for the F-dataset are shown in Fig. 3. RAFGAE achieves mean AUC and AUPR values of 0.9343 and 0.5270, respectively. The detailed results, including the F1-score and MCC, are presented in Table 3**.** The results based on the C-dataset are shown in Table 4. As shown in Tables 1 and 2, the newly proposed RAFGAE model

**Table 3** RAFGAE Model Performance via tenfold CV based on the F-Dataset

| Fold | AUC | AUPR | Precision | F1-score | Recall | MCC |
|---|---|---|---|---|---|---|
| 1 | 0.9230 | 0.5150 | 0.5843 | 0.5702 | 0.4934 | 0.6222 |
| 2 | 0.9410 | 0.5792 | 0.5529 | 0.6003 | 0.5539 | 0.6445 |
| 3 | 0.9289 | 0.5172 | 0.5538 | 0.5467 | 0.4742 | 0.5978 |
| 4 | 0.9290 | 0.4793 | 0.5779 | 0.5019 | 0.4059 | 0.5640 |
| 5 | 0.9362 | 0.5455 | 0.5684 | 0.5580 | 0.4835 | 0.6095 |
| 6 | 0.9415 | 0.5181 | 0.6298 | 0.5387 | 0.4344 | 0.6028 |
| 7 | 0.9392 | 0.5145 | 0.5592 | 0.5510 | 0.4779 | 0.6022 |
| 8 | 0.9369 | 0.5641 | 0.5462 | 0.5944 | 0.5487 | 0.6386 |
| 9 | 0.9335 | 0.5498 | 0.5422 | 0.5913 | 0.5461 | 0.6353 |
| 10 | 0.9337 | 0.4804 | 0.5830 | 0.5045 | 0.4075 | 0.5670 |
| Ave | 0.9343 ± 0.0006 | 0.5270 ± 0.0316 | 0.5697 ± 0.0026 | 0.5557 ± 0.0060 | 0.4825 ± 0.0025 | 0.6084 ± 0.0050 |

**Table 4** RAFGAE Model Performance via tenfold CV on the C-dataset

| Fold | AUC | AUPR | Precision | F1-score | Recall | MCC |
|---|---|---|---|---|---|---|
| 1 | 0.9396 | 0.5559 | 0.5712 | 0.6056 | 0.5517 | 0.6515 |
| 2 | 0.9431 | 0.5575 | 0.5737 | 0.6076 | 0.5532 | 0.6536 |
| 3 | 0.9446 | 0.4951 | 0.5538 | 0.5353 | 0.4541 | 0.5899 |
| 4 | 0.9489 | 0.5339 | 0.5848 | 0.5512 | 0.4671 | 0.6063 |
| 5 | 0.9431 | 0.5670 | 0.5611 | 0.5580 | 0.5438 | 0.6425 |
| 6 | 0.9452 | 0.4953 | 0.5663 | 0.5368 | 0.4553 | 0.5914 |
| 7 | 0.9388 | 0.4963 | 0.5663 | 0.5368 | 0.4553 | 0.5914 |
| 8 | 0.9495 | 0.5253 | 0.5911 | 0.5567 | 0.4719 | 0.6118 |
| 9 | 0.9434 | 0.5214 | 0.6242 | 0.5820 | 0.4924 | 0.6381 |
| 10 | 0.9496 | 0.5602 | 0.5654 | 0.6006 | 0.5473 | 0.6465 |
| Ave | 0.9446 ± 0.0006 | 0.5309 ± 0.0316 | 0.5757 ± 0.0025 | 0.5670 ± 0.0050 | 0.4992 ± 0.0033 | 0.6223 ± 0.0026 |

obtains good performance on the above two datasets, proving the effectiveness and robustness of this model.

### Parameter adjustment

Since the hyperparameter settings can influence the performance of RAFGAE, we used tenfold CV on the F-dataset to analyze the impact of different parameter settings. In the Re_GAT framework, the weight $\alpha$ of the initial residual connection and the weight $\beta$ of the adaptive residual connection can directly affect the result of feature fusion. To fully integrate adjacent node information and mitigate the oversmoothing problem, we adjust the $\alpha$ and $\beta$ values within the following range: $\alpha \in \{0.1 \sim 0.9\}$ and $\beta \in \{0.1 \sim 0.9\}$. As shown in Fig. 4, when $\alpha = 0.3$ and $\beta = 0.7$, the AUC reaches its maximum value.

In addition, the features of diseases and drugs are extracted via GATs. The Re_GAT framework computes and aggregates different multilayer features via the GAT. We discuss the impact of GATs with different numbers of layers on association prediction. Figure 5 presents the results of the ROC curve analysis on the basis of tenfold CV.

To optimize the initial parameters, we use the Adam optimizer [41]. As in previous studies [42, 43], we set the dropout and weight decay parameters to 0.5 and $10^{-5}$,

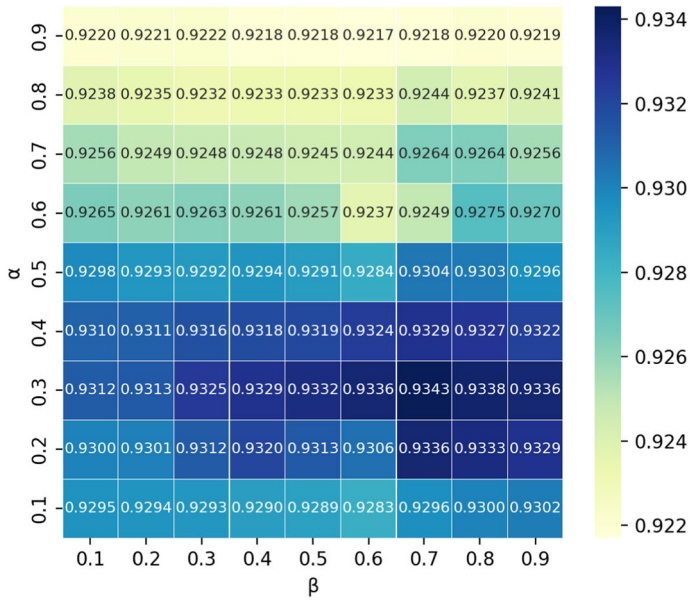


**Fig. 4** Effect of the *α* and *β* parameters on the AUC of RAFGAE

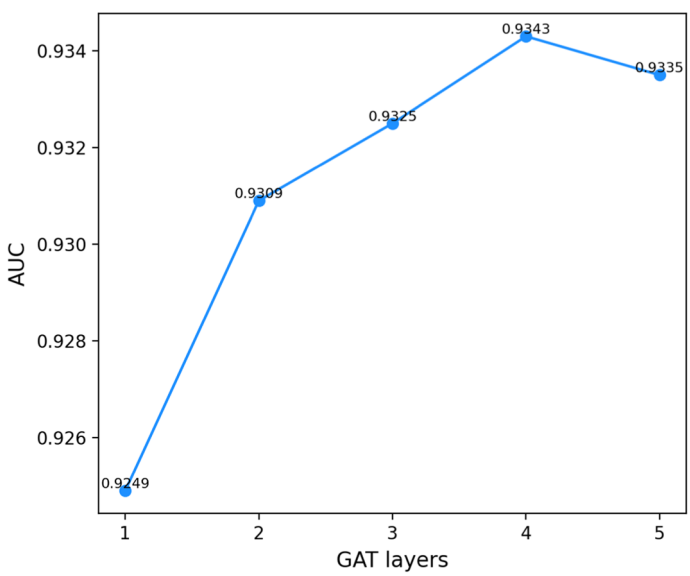


**Fig. 5** Effect of the number of GAT layers on the AUC of RAFGAE

respectively. We also evaluate the model performance by changing the dimensions of the GAE hidden layers. With the other parameters unchanged, the AUC value of RAFGAE generally increases as the embedding dimension of the GAE hidden layer increase and tends to stabilize when the dimension reaches 256. Finally, we set the embedding dimension of the hidden layer to 256. These results are shown in Fig. 6.

### Case studies

To evaluate the practical ability of RAFGAE to predict unknown indications of approved drugs as well as new therapies for existing diseases, we train the RAFGAE

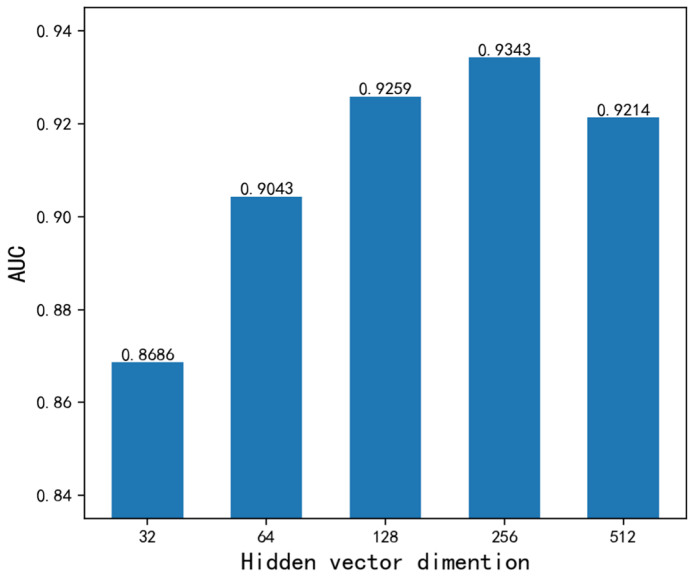


**Fig. 6** Effect of the hidden vector dimension on the AUC of RAFGAE

model using all known associations as training data, and predict potential associations for known diseases or drugs. The predicted ranking of unknown indications of approved drugs and unknown therapies for existing diseases is validated on the public database, namely, the Comparative Toxicogenomics Database (CTD) [44].

To assess the ability of RAFGAE to discover new indications, we select two representative medicinal products. Table 5 shows the confirmation information for the top 10 candidate diseases and the known drug-disease associations. Among them, doxorubicin is a cytotoxic anthracycline antibiotic that is widely used to treat various cancers, including Kaposi sarcoma and metastatic cancer related to AIDS. Of the top 10 positive predictions, there were 7 tumor-related diseases that have been verified via reliable databases. Levodopa is a precursor of dopamine and is commonly used in the treatment of Parkinson's syndrome and Parkinson's syndrome-related disorders because of its ability to cross the blood–brain barrier. As shown in Table 5, reliable sources have identified 7 of the top 10 associated diseases. This evidence suggests that RAFGAE can be trained on and can learn from existing biological information and can identify association markers that are not captured in the training set.

To validate the practical ability of RAFGAE to discover novel therapies, we select breast neoplasms and small-cell lung cancer as experimental cases. On the basis of the RAFGAE prediction results, the 10 drugs with the highest prediction scores are validated via the CTD. Table 6 shows similar results for the top 10 positive predictions. Breast neoplasms are among the most common malignancies in women and the leading cause of cancer-related disease in women. As shown in Table 6, 9 of the top 10 drugs were verified via reliable sources. The high incidence rate and high mortality of small cell lung cancer worldwide make this complex tumor a difficult medical problem. In summary, 6 drugs have been confirmed by evidence from authoritative sources among the top 10 predicted drugs ranked by prediction score. In summary, case studies have shown that RAFGAE can identify the associations between diseases

**Table 5** Top 10 related diseases for a given drug predicted by RAFGAE

| Drug (DrugBank ID) | Disease | OMIM ID | Evidence database |
|---|---|---|---|
| Doxorubicin (DB00997) | Esophageal cancer | 133,239 | CTD |
| | Small cell cancer of the lung | 182,280 | CTD |
| | Colorectal cancer | 114,500 | CTD |
| | Testicular germ cell tumor | 273,300 | CTD |
| | Leukemia | 109,543 | CTD |
| | Dohle bodies | 223,350 | N/A |
| | Reticulum cell sarcoma | 267,730 | CTD |
| | Renal failure | 161,900 | N/A |
| | Hypertension | 608,622 | CTD |
| | Mccune-albrightt syndrome | 174,800 | N/A |
| Levodopa (DB01235) | Parkinson disease | 168,600 | CTD |
| | Hyperplastic myelinopathy | 147,530 | CTD |
| | Mismatch repair cancer syndrome | 276,300 | N/A |
| | Asthma | 208,550 | CTD |
| | Growth retardation | 233,810 | CTD |
| | Gallbladder disease 1 | 600,803 | CTD |
| | Attention deficit | 143,465 | CTD |
| | Breast cancer | 114,480 | CTD |
| | Dohle bodies 1 | 223,350 | N/A |
| | Prolactin deficit | 264,120 | N/A |

**Table 6** Top 10 related drugs for a given disease predicted by RAFGAE

| Disease (OMIM ID) | Drug | DrugBank ID | Evidence database |
|---|---|---|---|
| Breast neoplasms (114,480) | Levamisole | DB00848 | CTD |
| | Diethylstilbestrol | DB00255 | CTD |
| | Adenosine phosphate | DB00131 | N/A |
| | Zoledronic acid | DB00399 | CTD |
| | Ascorbic acid | DB00126 | CTD |
| | Chlorthalidone | DB00310 | CTD |
| | Estramustine | DB01196 | CTD |
| | Atomoxetine | DB00289 | CTD |
| | Estrone | DB00655 | CTD |
| | Fenofibrate | DB01039 | CTD |
| Small cell lung cancer (182,280) | Doxorubicin | DB00997 | CTD |
| | Vincristine | DB00541 | CTD |
| | Carboplatin | DB00958 | CTD |
| | Isosorbide dinitrate | DB00883 | N/A |
| | Adenosine phosphate | DB00131 | N/A |
| | Fluorouracil | DB00544 | CTD |
| | Carmustine | DB00262 | CTD |
| | Dacarbaziner | DB00851 | N/A |
| | Leucovorin | DB00650 | N/A |
| | Mitomycin | DB00305 | CTD |

and drugs that are unknown in training datasets but that have been validated in real-world studies. Moreover, RAFGAE can make reliable predictions regarding unconfirmed potential associations between diseases and drugs. Therefore, RAFGAE has a noteworthy ability to uncover novel therapies/indications for existing diseases/drugs.

## Conclusions

In this paper, a deep-learning methodology named RAFGAE is developed for elucidating drug-disease associations. The key innovation of RAFGAE is that it combines the Re_GAT framework and the FMAT algorithm, facilitating the learning of neighbor node information and enhancing the initial node features in the disease-drug bipartite network. Then, two GAEs with collaborative training are applied to integrate the disease and drug spaces for association prediction. Notably, unlike some previous predictors that consider only low-order neighbor information, the Re_GAT framework can account for both high-order and low-order neighbor information by using multilayer GATs. Moreover, residual networks are introduced to mitigate model data oversmoothing, enabling the full employment of graph structure information hidden in the bipartite network. To enhance the initial features of nodes and make the model more robust, the FMAT algorithm is employed. This algorithm adds gradient-based adversarial perturbation to the input characteristics. In addition, we construct two GAEs with collaborative training for label propagation, enabling the full integration of the drug and disease space information for association prediction and improving the robustness of the RAFGAE model.

With tenfold CV, the RAFGAE model achieves an AUC score of 0.9343, which is better than the AUC scores of five state-of-the-art predictors. Furthermore, the case study results show that RAFGAE can reposition several representative drugs for human diseases and can be applied as a reasonable and effective tool for predicting the relationships between diseases and drugs.

We propose a computational drug repurposing method. This method can effectively identify candidate drugs with potential for treating different diseases and has the potential to uncover new indications for approved drugs that were previously unexplored. RAFGAE can guide wet laboratory experiments, accelerating drug development, reducing costs, and expanding treatment options. The method combines multilayer neural networks with residual connections to capture global information and alleviate oversmoothing problems. We also employ adversarial perturbations to improve the input quality. This novel combination of techniques provides a new perspective for future research and can also serve as a valuable reference for similar studies, such as predicting the associations between ncRNAs and diseases, microbiome-disease associations, and screening ncRNA drug targets.

However, RAFGAE has certain limitations. In this study, the negative and positive samples of the benchmark dataset are unbalanced, and we use all the negative samples as negative samples for training the proposed model. However, these unknown samples considered negative samples may be potential correlations, which greatly impacts the prediction accuracy of the model. In the future, we will select negative samples to further improve the model accuracy. In terms of biological data, we simply apply the interaction network between drugs and diseases without establishing a more informative biological regulatory network, which may further improve performance. In future research, we

will introduce other biological entities, such as proteins, pathways, and genes. In scenarios where drugs share the same or similar indications but lack structural similarity, the transmission of structural similarity information through a multilayer neural network can give rise to an "information leakage" problem, leading to a distorted view of the algorithm's performance in realistic drug repurposing settings. In our future research, we plan to address the problem of information leakage further by incorporating multiple drug similarities, such as target protein domain similarity, GO target protein annotation similarity, side effect similarity, and GIP similarity. This broader range of drug similarities can provide a more comprehensive features for drug repurposing. Similarly, incorporating disease similarities, such as disease ontology similarity, can help improve the accuracy and reliability of repositioning predictions by leveraging additional disease-related information.

## Methods

### Data preparation

We employ two benchmark datasets established by investigators. The first dataset is the F-dataset, which corresponds to Gottlieb's gold standard dataset [45]. The F-dataset contains 1933 known associations between diseases and drugs, including 313 diseases collected from the OMIM database [46] and 593 drugs obtained from the DrugBank database [47]. The second dataset is the C-dataset [24], which includes 2532 known associations between 409 diseases collected from the OMIM database and 663 drugs obtained from the DrugBank database. Table 7 summarizes the benchmark datasets in our proposal.

In this study, we calculated the drug structure similarity matrix $X_{dr}$ via the simplified molecular input line entry system (SMILES) chemical structure [48], which is represented as the Tanimoto index of chemical fingerprints of the drug pair via the Chemical Development Kit [49]. The disease semantic similarity matrix $X_{di}$ is computed from the semantic similarity of the disease phenotypes via information from the medical descriptions of the disease pairs [50].

### RAFGAE

After collecting the required data from different sources, we propose a prediction model with three individual modules to predict potential candidate diseases for drugs of interest. We first design the Re_GAT framework, which captures global structural information from a bipartite network. For the second module, we employ GAEs that use known associations between diseases and drugs to simulate label propagation to guide and predict unknown associations. On the basis of the above, we utilize the FMAT module for adversarial training

**Table 7** Details of the two benchmark datasets

| Dataset | No. of drugs | No. of diseases | No. of known associations |
|---|---|---|---|
| F-dataset | 593 | 313 | 1933 |
| C-dataset | 663 | 409 | 2532 |

to improve the input quality and increase the prediction accuracy. Figure 7 shows the overall workflow of RAFGAE.

### Re_GAT framework

Graph attention networks use a self-attention hidden layer to assign different attention scores to different neighbors, thus extracting the features of neighboring nodes more effectively.

The initial input to the Re_GAT framework can be described as follows:

$$h = \{h_1, h_2, \cdots h_N\}, h_i \in R^F \tag{1}$$

where $N$ represents the node count, $F$ represents the dimension of the feature and $h_i \in R^F$ represents the initial feature matrix of all the nodes. GATs calculate attention scores on the basis of the importance of neighbors and then aggregate neighbor features on the basis of the attention score.

The attention score is calculated as follows:

$$e_{ij} = \sigma\left(a^T\left[Wh_i \parallel Wh_j\right]\right) \tag{2}$$

To adjust for the influence of different nodes, we use the *softmax* function for attention score normalization score:

$$a_{ij} = soft\max_{j}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k\in N_i}\exp(e_{ik})} \tag{3}$$

By combining Formulas (3) and (4), the calculation formula for the attention score can be expressed as:

$$a_{ij} = \frac{\exp\left(\sigma\left(a^T[Wh_i \parallel Wh_j]\right)\right)}{\sum_{k\in N^i}\exp\left(\sigma\left(a^T[Wh_i \parallel Wh_k]\right)\right)} \tag{4}$$

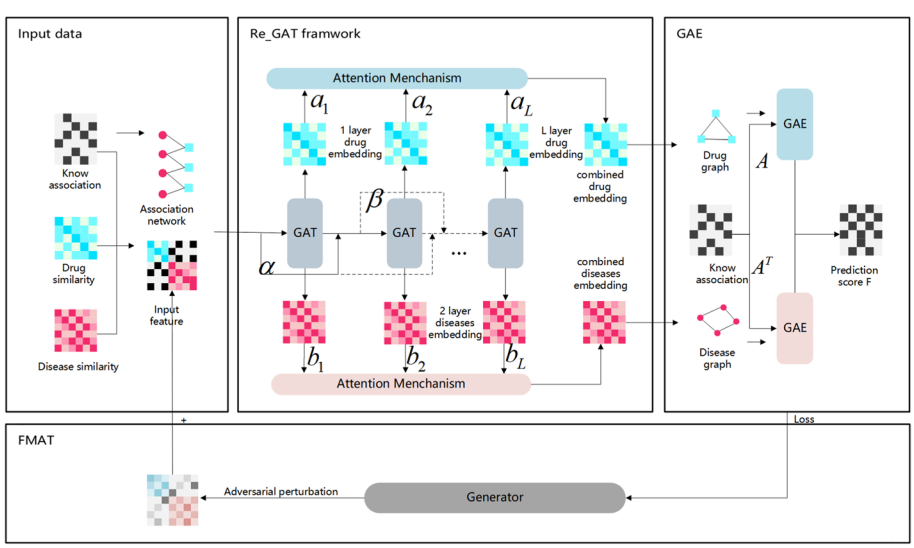


**Fig. 7** Flow chart of the RAFGAE calculation method

where $a_{ij}$ is the attention score, $W$ is a learnable linear transformation matrix, $a$ vector denotes the weight vector, $\sigma()$ represents the *LeakyReLU* activation function, and $\|$ denotes the connection operation. After normalization, the following formula can be used to calculate the final output feature:

$$h_i = \sigma\left(\sum_{\mathrm{j}\in N_i} a_{ij} W h_{\mathrm{j}}\right) \tag{5}$$

In this study, the drug-disease association matrix is given by matrix $A$, where the columns represent diseases and the rows represent drugs. The matrix $A(j, k) = 1$ if drug $j$ is associated with disease $k$ and 0 otherwise. Matrix $A$ and its transposition matrix $A^T$ define the bipartite network $G$:

$$G = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \in R^{(N_{\mathrm{dr}}+N_{\mathrm{di}})\times(N_{\mathrm{di}}+N_{\mathrm{dr}})} \tag{6}$$

We create the initial input embedding $H^{(0)}$ as follows:

$$H^{(0)} = \begin{bmatrix} X_{\mathrm{dr}} & A \\ A^T & X_{\mathrm{di}} \end{bmatrix} \in R^{(Ni+Nr)\times(Nr+Ni)} \tag{7}$$

When combined with the bipartite network adjacency matrix $G$ above, the graph attention network is defined as:

$$H^{(l)} = \sigma(GATs(G, H^{(l-1)})) \tag{8}$$

where $H^{(l)}$ represents the node embedding of the $l$-th layer, where $l = 1, \ldots, L$, and $GATs()$ represents a single attention layer, whereas the entire Re_GAT framework consists of multiple attention layers.

This study proposes a Re_GAT framework through two main strategies for forward propagation: (I) initial residual connection and adaptive residual connection; and (II) attention mechanism layer aggregation.

To facilitate the learning of feature information from higher-order neighbors, multiple attention layers are typically used, easily homogenizing the data and thus leading to oversmoothing problems. To alleviate the oversmoothing problem of deep CNNs, residual connections, also known as skip connections was first proposed for ResNet. Inspired by ResNet [51], recent studies have attempted to apply various residual connections to GATs to alleviate the oversmoothing problem. Several studies have shown that residual connections are necessary for deep GATs [52], not only to alleviate the oversmoothing problem, but also to give GATs a more stable gradient.

We sum the $H^{(l)}$ weights with $H^{(0)}$ and $H^{(l-1)}$ according to the scale coefficients $\alpha$ and $\beta$, respectively. We use the initial skip connection and the adaptive skip connection to mitigate the oversmoothing problem and accelerate the convergence of the GATs. The GAT formula of our model can be rewritten as:

$$H^{(l)} = \sigma\left(GATs(G, H^{(l-1)})\right) + \alpha H^{(0)} + \beta H^{(l-1)} \tag{9}$$

where $\alpha$ and $\beta$ are hyperparameters.

Inspired by LAGCN [35], the embedding of each layer captures structural information from different orders of the heterogeneous network. For instance, the initial layer obtains direct connection information, whereas the higher-order layers collect information about multihop neighbors through iterative update embedding. To fuse all useful information from multiple GAT layers, we use the attention mechanism. Since the Re_GAT framework calculates the embedding of different layers and the embeddings contain different information, we define the resulting GAT layer embedding as:

$$H_l = \begin{bmatrix} H_l^{dr} \\ H_l^{di} \end{bmatrix} \in R^{(Ndr+Ndi)\times kl} \tag{10}$$

where $Hdr\ l \in R^{Ndr\times kl}$ is the embedding of the drug in layer $l$ and $Hdi\ l \in R^{Ndi\times kl}$ is the embedding of the disease in layer $l$. We use attention mechanism layer aggregation to integrate multiple embedding matrices, and the final fused embedding matrix is as follows:

$$C_{dr} = \sum_{i=1}^{L} a_i H_i^{dr} \tag{11}$$

$$C_{di} = \sum_{i=1}^{L} \mathrm{b}_i H_i^{di} \tag{12}$$

where, $Hdr\ i$ and $Hdi\ i$ are the $l$-layer embeddings of drugs and diseases, respectively, $a_i$ and $b_i$ are the attention factors that can be calculated via Formulas (2), (3) and (4), and $L$ is the number of layers.

#### Constructing the feature similarity graph

A previous study showed that a similarity graph constructed using drug and disease features can be used to propagate labels [53]. We use the features $C_{dr}$ and $C_{di}$ to construct feature similarity graphs for diseases and drugs, respectively. These features are used for label propagation in the disease and drug spaces. The feature similarity graphs are constructed as follows. First, the Euclidean distance between nodes is calculated and ranked. Second, for each node $i$, its 10 nearest neighbors are selected. Finally, the adjacency matrix is defined as $M$, and the set of neighbors of node $i$ is defined as $N(i)$. The matrix $M$ satisfies $M_{ij}=1$ when $j$ belongs to $N(i)$; otherwise, $M_{ij}=0$.

The self-loop adjacency matrix for the similarity graph $S$ is constructed as follows:

$$S = M^T \odot M + I \tag{13}$$

where $\odot$ is the Hadamard product. This method can be used to obtain both the drug similarity graph $S_{dr}$ and the disease similarity graph $S_{di}$.

#### Graph autoencoder

Previous studies have shown that the graph autoencoder may simulate label propagation by iteratively propagating label information on the graph [54–56]. The association matrix $A$ can be considered initial label information. The initial label information

and the similarity graph $S$ calculated via the above method are input to the GAE. The encoder layer produces a hidden layer $Z$, whereas the decoder outputs the score $F$. The encoder of the GAE can be defined as:

$$GAE_{enc}(S, A) = \tanh(S \cdot ReLu(SA\Phi^{(0)})\Phi^{(1)}) \tag{14}$$

where $\Phi$ denotes the weight matrix. Here, we use two GAEs to propagate label information on the drug and disease graphs. We can obtain the drug hidden layer $Z_{dr}$ and the disease hidden layer $Z_{di}$, which are expressed as follows:

$$Z_{dr} = GAE_{enc}(S_{dr}, A) \tag{15}$$

$$Z_{di} = GAE_{enc}(S_{di}, A^T) \tag{16}$$

where $S_{dr}$ and $S_{di}$ denote the drug similarity graph and the disease similarity graph, respectively, and A denotes the association matrix.

The decoder of the GAE is applied to decode the hidden layer representation, which is defined as follows:

$$GAE_{dec}(S, Z) = sigmoid(S \cdot \mathrm{Re}Lu(SZ\Phi^{(2)})\Phi^{(3)}) \tag{17}$$

Therefore, the score matrices $F_{dr}$ and $F_{di}$ can be obtained by decoding $Z_{dr}$ and $Z_{di}$, respectively.

$$F_{dr} = GAE_{dec}(S_{dr}, Z_{dr}) \tag{18}$$

$$F_{di} = GAE_{dec}(S_{di}, Z_{di}) \tag{19}$$

Since $F_{dr}$ and $F_{di}$ are both low rank matrices [57], they need to satisfy the rank-sum inequality:

$$rank(\alpha F_{dr} + (1-\alpha)F_{di}^T) \leq rank(F_{dr}) + rank(F_{di}^T) \tag{20}$$

By performing a linear combination of $F_{dr}$ and $F_{di}$, the final integrated score is obtained as follows:

$$F = \alpha F_{dr} + (1-\alpha)F_{di}^T \tag{21}$$

where $\alpha \in (0,1)$ represents the balanced weight between the drug space and the disease space.

The GAE reconstruction error is the loss of cross-entropy between the final prediction and the true value:

$$L_r = -\sum_{i,j} A_{ij} \log F_{ij} \tag{22}$$

As the information from the disease space and the drug space influences the predicted outcome, we use a cotraining approach to train the above two GAEs. The cotraining training loss $L_{co}$ is defined as:

$$L_{co} = \frac{1}{2}\left\|Z_{dr}Z_{di}^{T} - A\right\|_{F}^{2} \tag{23}$$

The combined loss function can be rewritten as:

$$L = \alpha L_{rdr} + (1-\alpha)L_{rdi} + L_{co} \tag{24}$$

where $L_{rdr}$ and $L_{rdi}$ denote the reconstruction errors of the two GAEs in the drug space and the disease space, respectively.

#### Free multiscale adversarial training

In this section, we investigate how to effectively improve the input quality through data augmentation [58]. When neural networks are trained, the quality of the data is far more important than the quantity. By searching for and stamping out small perturbations that cause the classifier to fail, one may hope that adversarial training could benefit standard accuracy. Adversarial training is a well-studied method that increases the robustness and interpretability of neural networks. When the data distribution is sparse and discrete, the beneficial effect of adversarial perturbations on generalizability is prominent [59]. Inspired by this, we introduce free multiscale adversarial training (FMAT) to augment the node features [60].

Adversarial training first generates adversarial perturbations, which are then integrated into the training node features. Given a learning model $f_\theta$ with parameters $\theta$, we denote the perturbed feature as $H_{adv}=H+\delta$. Adversarial learning follows the min–max formulation:

$$\min_{\theta} E_{(H,A)\sim D}[\max_{\|\delta\|_p \le \varepsilon} L(f_\theta(H+\delta), A)] \tag{25}$$

where $A$ represents the real value, $D$ represents the data distribution, $L$ represents the objective loss function, $\varepsilon$ represents the perturbation budget, and $\| \ \|_p$ represents an $l_p$-norm distance measure.

The saddle-point optimization problem can be solved via projected gradient descent (PGD), which implements inner maximization, and stochastic gradient descent (SGD), which implements outer minimization. The parameter $\delta$ is updated after each step:

$$\delta_{t+1} = \prod_{\|\delta\|_\infty \le \varepsilon}(\delta_t + e \cdot sign(\nabla_\delta L(f_\theta(H+\delta_t), A))) \tag{26}$$

where $\prod_{\|\delta\| \le \varepsilon}$ is projected onto the $\varepsilon$-sphere under the $l_\infty$-*norm*. The initial layer of the Re_GAT framework can be rewritten as:

$$H^{(1)} = \sigma(GAT(H^{(0)} + \delta_t, G)) \tag{27}$$

To effectively exploit the generalizability of adversarial perturbations and improve their diversity and quality, Chen et al. emphasized the importance of adapting to different types of data enhancements [61]. To achieve this, we introduce a *'free'* training approach [62].

The calculation of $\delta$ is inefficient because the $N$-step update requires $N$ forward and backward channels. This update runs $N$ times completely forward and backward to obtain the worst perturbation $\delta_N$. However, the model weight $\theta$ is updated once to use

only $\delta_N$. Model training is $N$ times slower because of this process. In contrast, the *'free'* training outputs the model weights $\theta$ on the same backward channel while calculating the $\delta$ gradient, allowing model weight updates to be calculated in parallel with perturbation updates.

*'Free'* training has the same robustness and accuracy as standard adversarial training does. However, the training costs are the same as those of clean training. The *'free'* strategy accumulates a gradient of $\nabla_\theta L$ in each iteration and updates the model weight $\theta$ through this gradient. During training process, the model runs the inner circle $T$ times, each time calculating the gradient of $\theta_{t\text{-}1}$ and $\delta_t$ by taking a step along the average gradient at $H^{(l)} + \delta_0, \ldots, H^{(l)} + \delta_{T\text{-}1}$. Formally, the optimization step is

$$\min_\theta E_{(H,A)\sim D}\left[\frac{1}{T}\sum_{t=0}^{T-1}\max_{\|\delta\|_p\le\varepsilon} L(f_\theta(H+\delta_t),A)\right] \tag{28}$$

**Abbreviations**

| | |
|---|---|
| GAT | Graph attention network |
| GAE | Graph autoencoder |
| FMAT | Free multiscale adversarial training |
| TPR | True positive rate |
| FPR | False-positive rate |
| ROC | Receiver operating characteristic |
| AUC | Area under ROC curve |
| CV | Cross validation |

**Acknowledgements**
Not applicable.

**Author contributions**
GL and JL conceived and designed the study. GL and SL implemented the experiments and drafted the manuscript. CL and QX analyzed the results. All the authors have read and approved the final manuscript.

**Funding**
This work is supported by the National Natural Science Foundation of China (Grant Nos. 62362034, 61862025, 62372279, and 62002116), the Natural Science Foundation of Jiangxi Province (Grant Nos. 20232ACB202010, 20212BAB202009, 20181BAB211016), and the Natural Science Foundation of Shandong Province (Grant No. ZR2023MF119).

**Availability of data and materials**
We acquired the C-dataset of disease-drug associations, from the Comparative Toxicogenomics Database [44] (http://ctdbase.org/). We screened the F-dataset of disease-drug interactions from the OMIM database [46] (https://www.omim.org/) and DrugBank database [47] (https://www.drugbank.ca/). These two datasets and the source code are available at: https://github.com/ghli16/RAFGAE.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare no competing interests.

Received: 30 June 2023 Accepted: 6 August 2024
Published online: 08 August 2024

**References**

1. Rifaioglu AS, Atas H, Martin MJ, Cetin-Atalay R, Atalay V, Doğan T. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Brief Bioinform. 2019;20(5):1878–912.

2. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004;3(8):673–83.
3. Dickson M, Gagnon JP. Key factors in the rising cost of new drug discovery and development. Nat Rev Drug Discov. 2004;3(5):417–29.
4. Padhy BM, Gupta YK. Drug repositioning: re-investigating existing drugs for new therapeutic indications. J Postgrad Med. 2011;57(2):153.
5. Xue H, Li J, Xie H, Wang Y. Review of drug repositioning approaches and resources. Int J Biol Sci. 2018;14(10):1232.
6. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guilliams T, Latimer J, McNamee C, Norris A, Sanseau P, Cavalla C, Pirmohamed M. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41–58.
7. Baker NC, Ekins S, Williams AJ, Tropsha A. A bibliometric review of drug repurposing. Drug Discov Today. 2018;23(3):661–72.
8. Nosengo N. New tricks for old drugs. Nature. 2016;534(7607):314–6.
9. Jarada TN, Rokne JG, Alhajj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. J Cheminform. 2020;12(1):1–23.
10. Mohamed K, Yazdanpanah N, Saghazadeh A, Rezaei N. Computational drug discovery and repurposing for the treatment of COVID-19: a systematic review. Bioorg Chem. 2021;106: 104490.
11. Fahimian G, Zahiri J, Arab SS, Sajedi RH. RepCOOL: computational drug repositioning via integrating heterogeneous biological networks. J Transl Med. 2020;18(1):1–10.
12. Traylor JI, Sheppard HE, Ravikumar V, Breshears J, Raza SM, Lin CY, Patel SR, DeMonte F. Computational drug repositioning identifies potentially active therapies for chordoma. Neurosurgery. 2021;88(2):428.
13. Bai L, Scott MK, Steinberg E, Kalesinskas L, Habtezion A, Shah NH, Khatri P. Computational drug repositioning of atorvastatin for ulcerative colitis. J Am Med Inform Assoc. 2021;28(11):2325–35.
14. Dai W, Liu X, Gao Y, Chen L, Song J, Chen D, Gao K, Jiang YS, Yang YP, Chen JX, Lu P. Matrix factorization-based prediction of novel drug indications by integrating genomic space. Comput Math Methods Med. 2015;2015:275045.
15. Zhang W, Zou H, Luo L, Liu Q, Wu W, Xiao W. Predicting potential side effects of drugs by recommender methods and ensemble learning. Neurocomputing. 2016;173:979–87.
16. Huang F, Qiu Y, Li Q, Liu S, Ni F. Predicting drug-disease associations via multi-task learning based on collective matrix factorization. Front Bioeng Biotechnol. 2020;8:218.
17. Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. Bioinformatics. 2018;34(11):1904–12.
18. Zhang W, Yue X, Lin W, Wu W, Liu R, Huang F, Liu F. Predicting drug-disease associations by using similarity constrained matrix factorization. BMC Bioinform. 2018;19:1–12.
19. Yang M, Wu G, Zhao Q, Li Y, Wang J. Computational drug repositioning based on multi-similarities bilinear matrix factorization. Brief Bioinform. 2021;22(4):bbaa267.
20. Zhang W, Xu H, Li X, Gao Q, Wang L. DRIMC: an improved drug repositioning approach using Bayesian inductive matrix completion. Bioinformatics. 2020;36(9):2839–47.
21. Hu L, Zhang J, Pan X, Yan H, You ZH. HiSCF: leveraging higher-order structures for clustering analysis in biological networks. Bioinformatics. 2021;37(4):542–50.
22. Chu Y, Kaushik AC, Wang X, Wang W, Zhang Y, Shan X, Salahub DR, Xiong Y, Wei DQ. DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. Brief Bioinform. 2021;22(1):451–62.
23. Yang K, Zhao X, Waxman D, Zhao XM. Predicting drug-disease associations with heterogeneous network embedding. Chaos Interdiscip J Nonlinear Sci. 2019;29(12):123109.
24. Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, Peng J, Chen L, Zeng J. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nat Commun. 2017;8(1):573.
25. Zhao BW, Hu L, You ZH, Wang L, Su XR. HINGRL: predicting drug–disease associations with graph representation learning on heterogeneous information networks. Brief Bioinform. 2022;23(1):bbab515.
26. Zhang H, Cui H, Zhang T, Cao Y, Xuan P. Learning multi-scale heterogenous network topologies and various pairwise attributes for drug–disease association prediction. Brief Bioinform. 2022;23(2):bbac009.
27. Luo H, Wang J, Li M, Luo J, Peng X, Wu FX, Pan Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. Bioinformatics. 2016;32(17):2664–71.
28. Cai L, Lu C, Xu J, Meng Y, Wang P, Fu X, Su Y. Drug repositioning based on the heterogeneous information fusion graph convolutional network. Brief Bioinform. 2021;22(6):bbab319.
29. Xuan P, Ye Y, Zhang T, Zhao L, Sun C. Convolutional neural network and bidirectional long short-term memory-based method for predicting drug–disease associations. Cells. 2019;8(7):705.
30. Liu H, Zhang W, Song Y, Deng L, Zhou S. HNet-DNN: inferring new drug–disease associations with deep neural network based on heterogeneous network features. J Chem Inf Model. 2020;60(4):2367–76.
31. Peng L, Tan J, Xiong W, Zhang L, Wang Z, Yuan R, Li Z, Chen X. Deciphering ligand–receptor-mediated intercellular communication based on ensemble deep learning and the joint scoring strategy from single-cell transcriptomic data. Comput Biol Med. 2023;2023: 107137.
32. Xuan P, Gao L, Sheng N, Zhang T, Nakaguchi T. Graph convolutional autoencoder and fully-connected autoencoder with attention mechanism based method for predicting drug-disease associations. IEEE J Biomed Health Inform. 2020;25(5):1793–804.
33. Coşkun M, Koyutürk M. Node similarity-based graph convolution for link prediction in biological networks. Bioinformatics. 2021;37(23):4501–8.
34. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F. deepDR: a network-based deep learning approach to in silico drug repositioning. Bioinformatics. 2019;35(24):5191–8.
35. Yu Z, Huang F, Zhao X, Xiao W, Zhang W. Predicting drug–disease associations through layer attention graph convolutional network. Brief Bioinform. 2021;22(4):bbaa243.

36. Feng Q, Dueva E, Cherkasov A, Ester M. PADME: a deep learning-based framework for drug–target interaction prediction. https://arxiv.org/abs/1807.09741 (2019).
37. Meng Y, Lu C, Jin M, Xu J, Zeng X, Yang J. A weighted bilinear neural collaborative filtering approach for drug repositioning. Brief Bioinform. 2022;23(2):bbab581.
38. Gu Y, Zheng S, Yin Q, Jiang R, Li J. REDDA: integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction. Comput Biol Med. 2022;150: 106127.
39. Yang M, Luo H, Li Y, et al. Drug repositioning based on bounded nuclear norm regularization. Bioinformatics. 2019;35(14):i455–63.
40. Li J, Zhang S, Liu T, et al. Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. Bioinformatics. 2020;36(8):2538–46.
41. Kingma DP. A method for stochastic optimization. ArXiv Prepr. 2014.
42. Niu M, Zou Q, Wang C. GMNN2CD: identification of circRNA–disease associations based on variational inference and graph Markov neural networks. Bioinformatics. 2022;38(8):2246–53.
43. Shi Z, Zhang H, Jin C, Quan X, Yin Y. A representation learning model based on variational inference and graph autoencoder for predicting lncRNA-disease associations. BMC Bioinform. 2021;22(1):1–20.
44. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Rosenstein MC, Wiegers TC, Mattingly CJ. The comparative toxicogenomics database: update 2013. Nucleic Acids Res. 2013;41(D1):D1104–14.
45. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol. 2011;7(1):496.
46. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34(suppl_1):D668–72.
47. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33(suppl_1):D514–7.
48. Vidal D, Thormann M, Pons M. LINGO, an efficient holographic text based method to calculate biophysical properties and intermolecular similarities. J Chem Inf Model. 2005;45(2):386–93.
49. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The Chemistry Development Kit (CDK): an open-source Java library for chemo-and bioinformatics. J Chem Inf Comput Sci. 2003;43(2):493–500.
50. Van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA. A text-mining analysis of the human phenome. Eur J Hum Genet. 2006;14(5):535–42.
51. Kaiming H, Shaoqing R, Jian S. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770–778.
52. Sharma V, Dyreson C. Covid-19 screening using residual attention network an artificial intelligence approach. 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE. 2020:1354–1361.
53. Belkin M, Niyogi P, Sindhwani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. J Mach Learn Res. 2006;7(11).
54. Kipf TN, Welling M. Variational graph auto-encoders. https://arxiv.org/abs/1611.07308 (2016).
55. Li G, Luo J, Xiao Q, Liang C, Ding P. Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity. J Biomed Inform. 2018;82:169–77.
56. Wang F, Zhang C. Label propagation through linear neighborhoods. Proceedings of the 23rd international conference on Machine learning. 2006:985–992.
57. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. https://arxiv.org/abs/1409.0473 (2014).
58. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.
59. Gan Z, Chen YC, Li L, et al. Large-scale adversarial training for vision-and-language representation learning. Adv Neural Inf Process Syst. 2020;33:6616–28.
60. Kong K, Li G, Ding M, Wu Z, Zhu C, Ghanem B, Taylor G, Goldstein T. Robust optimization as data augmentation for large-scale graphs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:60–69.
61. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In International conference on machine learning. PMLR. 2020:1597–1607.
62. Shafahi A, Najibi M, Ghiasi MA, Xu Z, Dickerson J, Studer C, Davis LS, Taylor G, Goldstein T. Adversarial training for free!. Adv Neural Inf Process Syst. 2019;32.

## Publisher's Note