iScience

CellPress
OPEN ACCESS

Article

Multi-task learning for predicting synergistic drug combinations based on auto-encoding multi-relational graphs

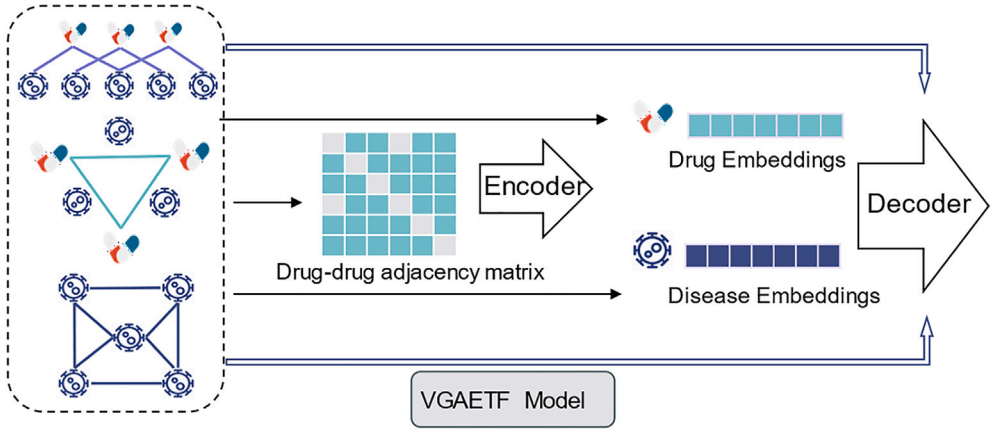
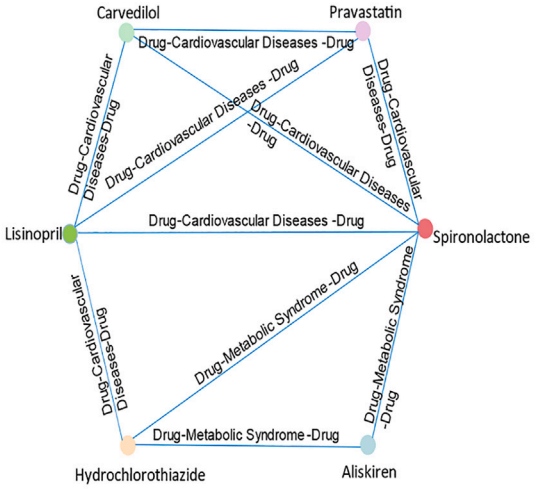


Table 3. Top 10 predicted hydrochlorothiazide -containing drug combinations for treating hypertension

Rank	Drug Combination (triplets)	Evidence Source
1	Hydrochlorothiazide -hypertension - Spironolactone	PMID: 363334
2	Hydrochlorothiazide -hypertension - Lisinopril	PMID: 27703294
3	Hydrochlorothiazide -hypertension - Labetalol	PMID: 2860990
4	Hydrochlorothiazide -hypertension - Aliskiren	PMID: 21122057
5	Hydrochlorothiazide -hypertension - Carvedilol	PMID: 1974505
6	Hydrochlorothiazide -hypertension - Valsartan	PMID: 11800062
7	Hydrochlorothiazide -hypertension - Eplerenone	PMID: 27651039
8	Hydrochlorothiazide -hypertension - Raltegravir	N/A
9	Hydrochlorothiazide -hypertension - Amlodipine	PMID: 19470877
10	Hydrochlorothiazide -hypertension - Colchicine	N/A



Visualization of the subgraph for explaining the spironolactone central node

Wenyu Shan, Cong Shen, Lingyun Luo, Pingjian Ding

dpj@usc.edu.cn

Highlights

An auto-encoding-based computational method for synergistic drug combination prediction

Construct a multi-relational graph to model complex relationships between drugs

We incorporate multi-task learning to improve the VGAETF's generalization ability

Shan et al., iScience 26, 108020
October 20, 2023
https://doi.org/10.1016/j.isci.2023.108020

iScience

CellPress
OPEN ACCESS

Article

Multi-task learning for predicting synergistic drug combinations based on auto-encoding multi-relational graphs

Wenyu Shan,[1] Cong Shen,[2] Lingyun Luo,[1,3] and Pingjian Ding[1,4,*]

SUMMARY

Combinatorial drug therapy is a promising approach for treating complex diseases by combining drugs with synergistic effects. However, predicting effective drug combinations is challenging due to the complexity of biological systems and the limited understanding of pathophysiological mechanisms and drug targets. In this paper, we proposed a computational framework called VGAETF (Variational Graph Autoencoder Tensor Decomposition), which leveraged multi-relational graph to model complex relationships between entities in biological systems and predicted disease-related synergistic drug combinations in an end-to-end manner. In the computational experiments, VGAETF achieved high performances (AUROC [the area under receiver operating characteristic] = 0.9767, AUPR [the area under precision-recall] = 0.9660), outperforming other compared methods. Moreover, case studies further demonstrated the effectiveness of VGAETF in identifying potential disease-related synergistic drug combinations.

INTRODUCTION

Combinatorial drug therapy involves the simultaneous use of two or more compounds to treat a disease, potentially improving therapeutic efficacy through synergistic effects. This approach has gained significant attention and is increasingly utilized in the management and treatment of sophisticated diseases such as cancer, HIV/AIDS,[1] Hepatitis C,[2] and hypertension.[3] Combination therapies offer hope for treating diseases with unmet medical needs. Additionally, drug combinations often involve the use of already approved drugs by the Food and Drug Administration (FDA), which have well-studied toxic properties and side effects. Therefore, patients can safely use these combinations.[4] Experimental identification of drug combinations with synergistic effects is a time-consuming and costly process, made more difficult by the large amount of potential drug combinations.

Computational methods have emerged as an essential tool for predicting and prioritizing pairs of synergistic drugs. Machine learning algorithms are commonly used to mine potential combination pharmacotherapy from vast amounts of data, providing guidance for clinical trials and reducing costs.[5] However, it is challenging to extract disease-specific pathophysiological mechanisms and identify drug targets from clinical and laboratory data to predict drug combinations.[6] To address this issue, researchers have developed innovative approaches that incorporate diverse biological data. For instance, Jiang et al.[7] predicted synergistic drug combinations for anticancer treatment by building a heterogeneous graph that involves drugs and proteins. Cheng et al.[8] predicted forecast disease-drug-drug combinations by quantifying the correlations between disease proteins and drug targets in protein interactomes. Li et al.[9] employed large pre-trained language models, such as generative pre-trained transformer (GPT), to predict the synergistic effects of drug pairs in a few-shot learning scenario. Zhang et al.[10] developed a comprehensive multi-relational graph that incorporates drugs, targets, and enzymes to capture intricate semantic information from diverse entities and utilized drug and cell line embeddings to predict the synergy score of drug combinations. Wu et al.[11] improved the deep forest-based method that predicts synergistic drug combinations in cancer cell lines. However, existing prediction approaches rely heavily on pharmacological responses and a thorough comprehension of the genetic and molecular mechanisms of disease, which are often restrictive.[12]

Recently, the embedding of multi-relational graphs has been a valuable tool for drug discovery.[13] Different from traditional graphs and networks that contain only one type of relationship, the multi-relational graph provides heterogeneous information that comprises multiple entities (e.g., proteins, drugs, and targets) and various types of relationships (e.g., drug-drug interactions or drug-target pairs).[14] Multi-relational graphs utilize nodes to represent entities and edges to represent the relationships between them, providing unstructured semantic relationships between entities.[15] As a result, multi-relational graphs are well-suited for modeling the intricate relationships that exist among entities within biological.[16] By representing biomedical entities as nodes and their relationships as edges in a graph structure, the

[1]School of Computer Science, University of South China, Hengyang, Hunan 421001, China
[2]College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan 410082, China
[3]Hunan Medical Big Data International Science and Technology Innovation Cooperation Base, Hengyang, Hunan 421001, China
[4]Lead contact
*Correspondence: dpj@usc.edu.cn
https://doi.org/10.1016/j.isci.2023.108020

Table 1. Statistics of datasets used in this study

Samples	Source	Number
Drugs	Shtar, G. et al.[26]	1105
Diseases	Shtar, G. et al.[26]	1283
Training data	Shtar, G. et al.[26]	53694
Testing data	Shtar, G. et al.[26]	5965
Drug-disease-drug triplets	Shtar, G. et al.[26]	59659
Drug-disease associations	Shtar, G. et al.[26]	26425
Disease-disease associations	Ioannidis, V. et al.[27]	259

multi-relational graph is able to learn lower-order representations of entities and relationships simultaneously. This preserves the inherent structure of the multi-relational graph.[17] The learned vectors are used to generate a score for each triplet by a scoring function that gives high scores to true triples and low scores to false ones. This way, the model can infer whether triplets are true or not. Multi-relational graph embedding models include translation-based models like TransE,[18] bilinear models like ComplEx,[19] CP,[20] TuckER,[21] neural network-based models, and graph convolutional networks (GCNs)-based models[22] including R-GCN,[23] ComplexGCN,[24] etc. Translation-based models treat the relation as a vector that moves the head entity close to the tail entity in an embedding space. The models then obtain a triplet score based on the distance. Bilinear models use multiplicative scoring functions to match the latent semantics of entities and relations, while neural network-based models employ neural network models to learn their interactions. Graph neural networks (GNNs) are used for learning connectivity structure, with the GCN[25] acting as a graph encoder.

In this study, we propose VGAETF, an effective framework for predicting disease-related drug combinations by seamlessly integrating drug-oriented and disease-oriented information using multi-relational graph neural networks. We incorporate an auxiliary task of reconstructing the drug-disease relationship matrix and disease-disease relationship matrix to improve the model's generalization ability. First, we construct a network that involves a drug-disease-drug multi-relational graph, a drug-disease association network, and a disease-disease association network. Second, we then use the variational graph autoencoder to implement a variational inference model that learns the probability distributions of low-dimensional vector representations of drugs. Third, we employ an inner product decoder to learn the representation matrix from disease-resembles-disease and drug-disease relationships and calculate the prediction scores of the triplet facts using tensor decomposition. VGAETF then learns drug and disease features under loss constraints to reveal the underlying structure of drug-disease-drug triplets. Our experimental results show that VGAETF outperforms other compared computational methods in drug combination detection. Furthermore, case studies investigating the effectiveness of VGAETF in predicting drug combinations serve as compelling evidence of its predictive prowess.

RESULTS

The framework of VGAETF

We obtained a dataset from Continuous Drug Combination Database (CDCDB),[26] a comprehensive drug combination database containing more than 40,795 drug combinations. We extracted the DrugBank identifiers (IDs) that are related to the FDA Orange Book and the Medical Subject Headings (MeSH) terms that describe the conditions that are treated in the study from our database. The data schema ($drug_i$, $disease_k$, $drug_j$) shows that combining $drug_i$ and $drug_j$ can treat $disease_k$. Among the labeled "Efficacious" drug pairs for one or more diseases, there are a total of 1,105 individual drugs, 1,283 diseases, and 4,709 pairs of drugs. We constructed 59,659 positive samples of drug-disease-drug triplets using this information. In this study, we randomly selected 90% of the positive samples as training data, while reserving the remaining 10% as testing data. Based on the drugs and diseases obtained from the training data, we obtained 26,425 drug-disease associations in the association data. We construct a binary matrix to denote known drug-disease associations in which values of 1 and 0 indicate verified associations and unverified associations between drugs and diseases, respectively. We collate 259 disease-disease associations by Ioannidis, V. et al.[27] This dataset contains disease-resembles-disease edges that are based on the co-occurrence of disease terms in PubMed. We also construct a binary matrix that includes two values 1 and 0, denoting validated and unvalidated associations between diseases and diseases, respectively. More details can be found in Table 1.

Figure 1 illustrates the structure of VGAETF which follows the popular encoder-decoder framework. We constructed the drug-disease-drug multi-relational graph, the drug-disease association network, and the disease-disease association network. The VGAETF encoder comprising an entity encoder uses the variational graph autoencoder[28,29] to learn the probability distributions of low-dimensional vector representations of drugs. To improve predictive performance, we use an inner product decoder to learn the representation matrix of disease-disease and drug-disease relationships. Including disease-disease relationships can reduce noise and missing data in triplets. Specifically, the learned drug and disease node features are reconstructed into a drug-disease relationship matrix and a disease-disease relationship matrix through an inner product decoder. Therefore, VGAETF employs multi-task learning to deal with the prediction problem of disease-related drug combinations as the primary task and the auxiliary task (i.e., learning disease-disease relationships and drug-disease relationships). The two tasks share the probability distribution of the low-dimensional vector representation learned by the encoder. At this step, VGAETF utilizes

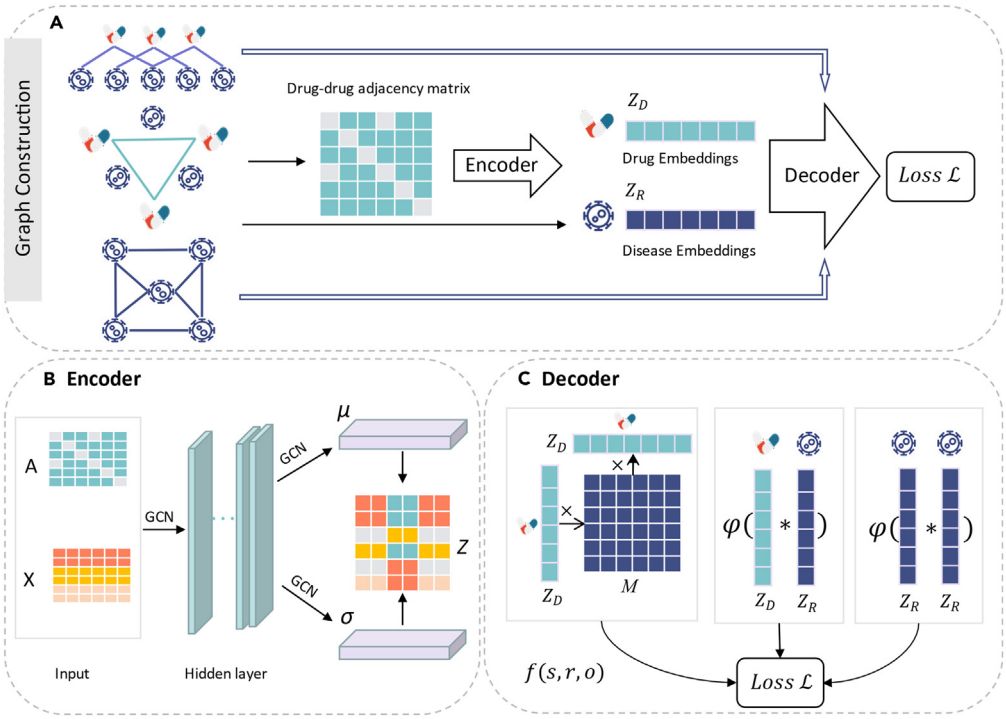


Figure 1. Overview of the VGAETF architecture

(A) Construct a drug-disease-drug multi-relational graph, drug-disease association network, and disease-disease association network. For each triplet fact (d_i, r_k, d_j), we get the corresponding embeddings z_i and z_j of the source node d_i and target node d_j from learned drug features. In addition, we set a trainable diagonal matrix M_k for each edge type k. VGAETF learns features for drugs and diseases under the constraint of losses.

(B) VGAETF uses a variational graph autoencoder to implement a variational inference model and learn the probability distributions of drugs Z_D.

(C) VGAETF employs the inner product as a decoder to reconstruct the drug-disease association network and the disease-disease association network from the learned latent features and adopts $f(s, r, o)$ to calculate the prediction scores of the triplet facts.

the encoder to diffuse the associations and node feature information among the association networks and learn features for drugs and diseases under the constraint of losses.

Experimental setting

We evaluated the performance of the VGAETF model using the area under the receiver operating characteristic (ROC) curve (AUC)[30] and the area under the precision/recall curve (AUPRC).[31] AUC is a binary classification evaluation metric commonly employed in multi-relational graph link prediction, measuring the ability of the model to differentiate between positive and negative samples by computing the area under the curve of the ROC curve. The area under precision-recall (AUPR) is another commonly used evaluation metric, which is especially useful when the positive class is rare or imbalanced, measuring the trade-off between precision and recall by computing the AUPRC. It provides a more informative evaluation of the model performance in the high-recall regime, which is typically more relevant in the context of link prediction in multi-relational graphs.

We implemented the VGAETF models with PyTorch[32] and DGL packages.[33] The publicly available source codes were used to implement all baseline methods. The best or default parameters of each method are used. The implementation of TransE, ComplEx, SimplE, and relational graph convolutional network (RGCN) models is provided by the Open Graph Benchmark.[34] We use the model implementation of TuckER, ComplexGCN, and knowledge graph embedding-based method for predicting drug combinations (KGE-DC) provided by the authors.

Hyperparameters selection

In this study, we utilized grid search to determine the optimal parameters for our model. Specifically, we initialized the disease embeddings using Xavier initialization[35] and optimized the training loss using the Adam solver.[36] Various hyperparameters were taken into account in our model. The hyperparameters in the VGAETF model, which include the learning rate, the number of units in hidden layer 1 and hidden layer 2, and the balance factor α, were considered. The hyperparameter α balances the contributions of the representation matrices to learn disease-disease relationships and drug-disease relationships in the auxiliary task. To prevent overfitting and improve efficiency, early stopping was implemented during training, and the number of training epochs was fixed at 1,200. The optimal hyperparameters were selected and highlighted in bold in Table 2.

Table 2. Hyperparameters settings considered for VGAETF, where bolded are the hyperparameters chosen for the model

Hyperparameter	Values considered
Learning rate	0.01; **0.005**; 0.001
The number of units in hidden layer 1 (h1)	64; 128; 192; 256; 320; **384**;448; 512
The number of units in hidden layer 2 (h2)	64; 128; 192; 256; **320**; 384; 448; 512
Balance factor α	0.1; 0.2; 0.3; 0.4; 0.5; **0.6**; 0.7; 0.8; 0.9

Embedding visualization

Our multi-relational graph embedding models are capable of automatically learning feature vectors for drugs and diseases. To verify the effectiveness of our model, we visualized the disease embedding vectors in a two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE).[37] However, due to the potential redundancy and lack of clarity that arises from visualizing and labeling all the nodes, we randomly selected two hundred disease nodes for visualization. By randomly selecting two hundred nodes, we aimed to maintain the representative nature of the dataset. The selection process ensured a diverse sampling from the entire range of diseases, providing an adequate representation of the overall distribution and characteristics. The resulting visualization, shown in Figure 2, reveals the relationships that exist between the embedding vectors of diseases. We observed that diseases that are closely embedded often co-occur in drug combinations, indicating that our model's learned disease embedding vectors can accurately capture the internal relational structure of diseases within the multi-relational graph of disease-related drug combinations.

For instance, after dimensionality reduction, we found that the diseases of Lung Neoplasms were closely related to Respiratory Tract Diseases, Thoracic Neoplasms, and Respiratory Tract Neoplasms within the two-dimensional space. These three diseases frequently occur in drug combinations together with lung tumors through known triplets. For example, the drug combination of Docetaxel and Pemetrexed, the drug combination of Pembrolizumab and Gemcitabine, and the drug combination of Pemetrexed and Atezolizumab are used to treat these diseases.[38–41] The findings suggest that our model can effectively manage the intricate relationships inherent in the multi-relational graph of drug combinations related to diseases.

Importance of each component

In this section, a series of studies were conducted to evaluate the contributions of different components to the VGAETF. Specifically, three variants of the VGAETF were implemented, including VGAETF w/o RR, VGAETF w/o DR, and VGAETF w/o DRR. In the VGAETF w/o RR variant, the auxiliary task of reconstructing disease-disease relationship networks was removed. In the VGAETF w/o DR variant, the auxiliary task of reconstructing the drug-disease relationship was removed. Lastly, in the VGAETF w/o DRR variant, both the auxiliary tasks of reconstructing disease-disease relationships and drug-disease relationships were eliminated. These modifications enabled the investigation of the effects of each auxiliary task on the VGAETF's performance. As shown in Figure 3, VGAETF consistently outperformed the three variants. Specifically, VGAETF w/o RR had slightly higher performance than VGAETF w/o DR with AUPR scores ranging from 0.9626 to 0.9615, while VGAETF w/o DRR had lower performances than VGAETF w/o RR and VGAETF w/o DR, with AUPR scores ranging from 0.9626 to 0.9601, and AUC scores ranging from 0.9746 to 0.9742. These results are consistent with our intuition that understanding the associations between drugs and diseases can provide more comprehensive biological information. Such associations provide valuable information for synergetic drug combination modeling, enabling more accurate prediction of the efficacy of drugs across different diseases and the discovery of new opportunities for drug repurposing, ultimately improving the predictive power of the model.

Performance evaluation

Currently, there exist several advanced multi-relational graph embedding techniques for the purpose of predicting triplets in bioinformatics. It should be noted that these models were not developed specifically for drug-disease-drug association prediction. We only used these algorithms for comparative analysis. The baselines we use for comparative analysis include translational distance-based techniques, semantic matching-based techniques, and neural network-based techniques. The details of each method are dicussed here. TransE[18] is a popular multi-relational graph embedding model that represents entities and relations as vectors, aiming to find embeddings that satisfy translation-based scoring functions between head and tail entities. ComplEx[19] extends TransE by representing entities and relations as complex vectors and using the Hermitian dot product to model both symmetric and antisymmetric relations. SimplE[42] is a simplified version of ComplEx that reduces the number of parameters by using diagonal matrices for relation embeddings and enforcing a sparsity constraint on them. TuckER[21] combines the Tucker decomposition[43] and bilinear forms to model high-order interactions between entities and relations. R-GCN[23] is a GCN that operates on a graph with labeled nodes and relations and utilizes relational graph convolutions to encode node features and learn representations of the graph, and RGCN has also been employed to predict drug combinations (drug-cell line-drug triplet).[7] ComplexGCN[24] extends RGCN by incorporating complex embeddings to handle both real and imaginary parts of relation weights and introduces a novel complex-valued graph convolution operation. KGE-DC[10] constructed a multi-relational graph and combined both drug chemical features and cell line genomic features to learn drug and cell line embeddings to predict the synergy score of drug combinations for cell lines.

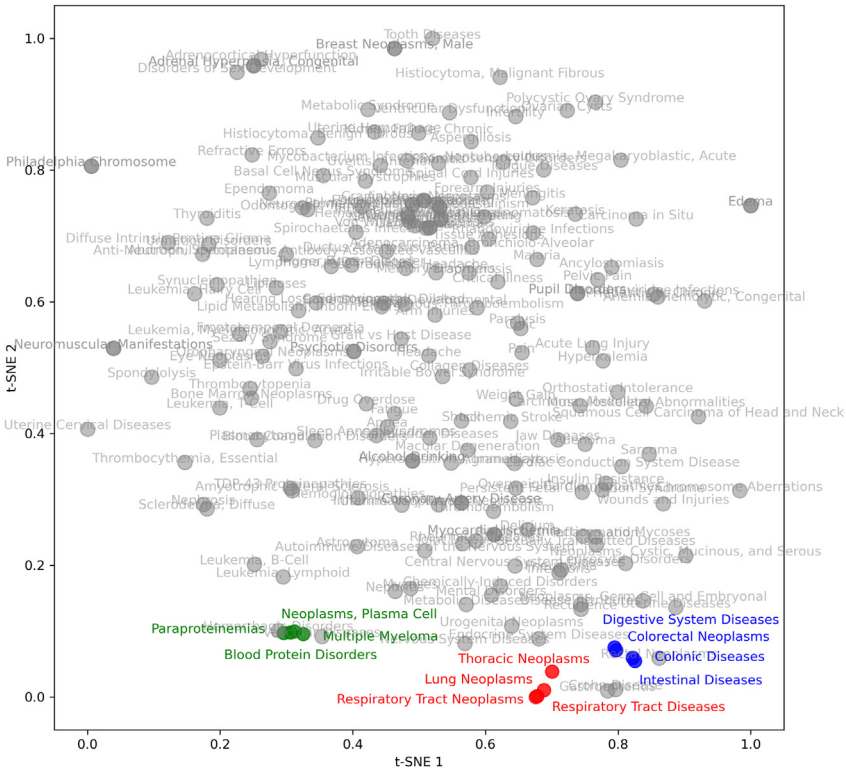


Figure 2. The visualization of disease embedding in VGAETF using the t-SNE package[37]

From this figure, we can observe that closely embedded diseases often co-occur in drug combinations, suggesting that VGAETF is able to learn disease interconnections in the multi-relational graph of disease-related drug combinations.

To assess the predictive capabilities of VGAETF in comparison with the baselines, a 5-fold cross-validation (CV) was conducted on a randomly selected 90% of the dataset. The remaining 10% of the samples were reserved as an independent test set to prevent overfitting and ensure a realistic assessment of the model's predictive accuracy. Both VGAETF and the baselines were evaluated on the independent test set to ensure unbiased results. Furthermore, an equivalent number of negative instances were randomly selected from the missing triplets to match the number of positive samples. The entire CV set was used to train all methods, which were subsequently used for predictions on the independent test set to enable objective evaluation.

We compared our model with the seven baseline methods, as depicted in Figure 4. Notably, the obtained results showed that the best AUC and AUPR values of 0.9767 and 0.9660 were achieved using VGAETF, outperforming other computational methods. The superior performance of VGAETF can be attributed to its utilization of the encoder-decoder operation of multi-relational embedding techniques, which incorporates the drug-disease-drug multi-relational graph structure, disease-disease relationship, and drug-disease relationships. The SimplE, KGE-DC, and ComplEx models exhibited remarkable performance, as indicated by their high scores on AUC and the AUPR. Specifically, the SimplE model achieved AUC and AUPR scores of 0.9671 and 0.9557, respectively, highlighting its ability to capture latent representations across various variables. Similarly, the ComplEx model leveraged complex space embeddings to enhance link prediction performance, achieving scores of 0.9397 and 0.9282 for AUC and AUPR, respectively. Furthermore, KGE-DC was instantiated through the integration of a multi-relational graph embedding technique and a neural network and attained an AUC of 0.9717 and an AUPR of 0.9559. These results underscore the efficacy of representing drug-disease-drug interactions as a multi-relational graph, enabling the acquisition of valuable topological insights. Overall, the proposed algorithm employing multi-relational graph embedding directly can yield relatively satisfactory results and exhibit excellent scalability when applied to large-scale datasets.

DISCUSSION

To further validate the reliability of VGAETF, we performed detailed case studies on computationally predicted candidate drugs for new anti-hypertension that contain hydrochlorothiazide and new anti-breast cancer that contain trastuzumab. The entire set of drug combinations was used to train the VGAETF model for predicting new anti-hypertension drug pairs that contain hydrochlorothiazide and new anti-breast cancer drug pairs that contain trastuzumab. We masked known triplets of anti-hypertension drug pairs containing hydrochlorothiazide and anti-breast cancer drug pairs containing trastuzumab in our samples, which did not appear in model training. The drug candidates were ranked

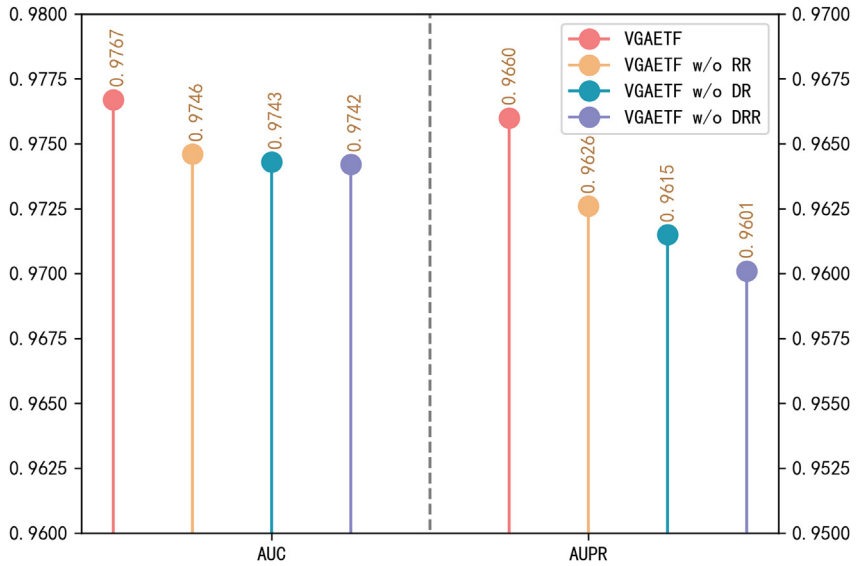


Figure 3. The VGAETF ablation experiment compared the performance of VGAETF with three different variants

VGAETF w/o RR did not try to reconstruct the auxiliary task of disease-disease relationships. VGAETF w/o DR did not try to reconstruct the auxiliary task of drug-disease relationships. VGAETF w/o DRR did not do either.

according to the VGAETF prediction scores for the triplets, and the top 10 scoring candidates are shown in Tables 3 and 4. The rankings of the triplets are based on the predicted scores calculated by the scoring function on embeddings of drugs and diseases, with higher scores corresponding to higher rankings. To indicate the presence of supporting literature, we used the PMID number in PubMed, and "N/A" indicates a lack of confirmation for the association. The results of case studies indicate that our model is indeed able to predict potential disease-related drug combinations.

Hydrochlorothiazide, a commonly used antihypertensive drug, is a diuretic that lowers blood pressure by reducing the amount of salt and water in the body.[44] The top 10 predicted candidates for hydrochlorothiazide-containing drug combinations for the treatment of hypertension are listed in Table 3, where eight candidates were supported by clinical studies, such as spironolactone, lisinopril, labetalol, aliskiren, carvedilol, valsartan, eplerenone, and amlodipine.

Hydrochlorothiazide-spironolactone is a potent diuretic combination that effectively treats hypertension by reducing water retention through increased urine output.[45] Hydrochlorothiazide-lisinopril can produce a synergistic effect, lowering blood pressure more effectively than either drug alone, and improving patient adherence.[46,47] Hydrochlorothiazide-labetalol can be used in combination to achieve better blood pressure control, and labetalol is an alpha-beta blocker that can lower blood pressure by reducing the force of the heart's contractions and heart rate.[48] Hydrochlorothiazide-aliskiren can achieve better blood pressure reduction than using either drug alone, especially in obese patients with hypertension, and aliskiren is a hormonal pathway that regulates blood pressure and fluid balance.[49,50] Hydrochlorothiazide-carvedilol can be combined to treat hypertension, and carvedilol is a beta-blocker that can lower heart rate and blood pressure.[51] The combination of hydrochlorothiazide-valsartan has been shown to be effective in the treatment of hypertension, and the advantage of this combination therapy is that it can reduce the dose, thereby reducing adverse reactions.[52,53] The combination of hydrochlorothiazide and eplerenone can produce a synergistic effect, resulting in more effective blood pressure reduction, and eplerenone is an aldosterone receptor antagonist that lowers systolic and diastolic blood pressure.[54] Combining hydrochlorothiazide with amlodipine has been demonstrated to be effective. Amlodipine, a calcium channel blocker, can lower blood pressure by dilating blood vessels and reducing vascular resistance and pressure.[55,56]

To elucidate the mechanism of action of our model in predicting novel drug combinations, we apply GNNExplainer,[57] a general tool for explaining GNN predictions. We selected the top predicted hydrochlorothiazide-containing drug combinations for treating hypertension in the case study. We used GNNExplainer to identify the most influential nodes in the trained model with respect to the spironolactone central node and analyzed their roles in the prediction. Figure 5 shows the results of the subgraph centered on spironolactone obtained by GNNExplainer. We found a combination of hydrochlorothiazide and spironolactone for treating metabolic syndrome. Additionally, lisinopril with hydrochlorothiazide or spironolactone could be combined to treat cardiovascular diseases (CVDs). Based on the literature on disease comorbidities, hypertension is a significant risk factor for various CVDs, including stroke and heart failure, with CVD being the complication of chronic kidney insufficiency (CKI).[58] The relationship between hypertension and CKI often coexists in patients, leading to a complex interplay between the two conditions.[59] Furthermore, the efficacy of hydrochlorothiazide in controlling hypertension has been established. Considering that metabolic syndrome and hypertension frequently coexist, these complex associations may suggest a potential benefit of hydrochlorothiazide and spironolactone in the treatment of hypertension.

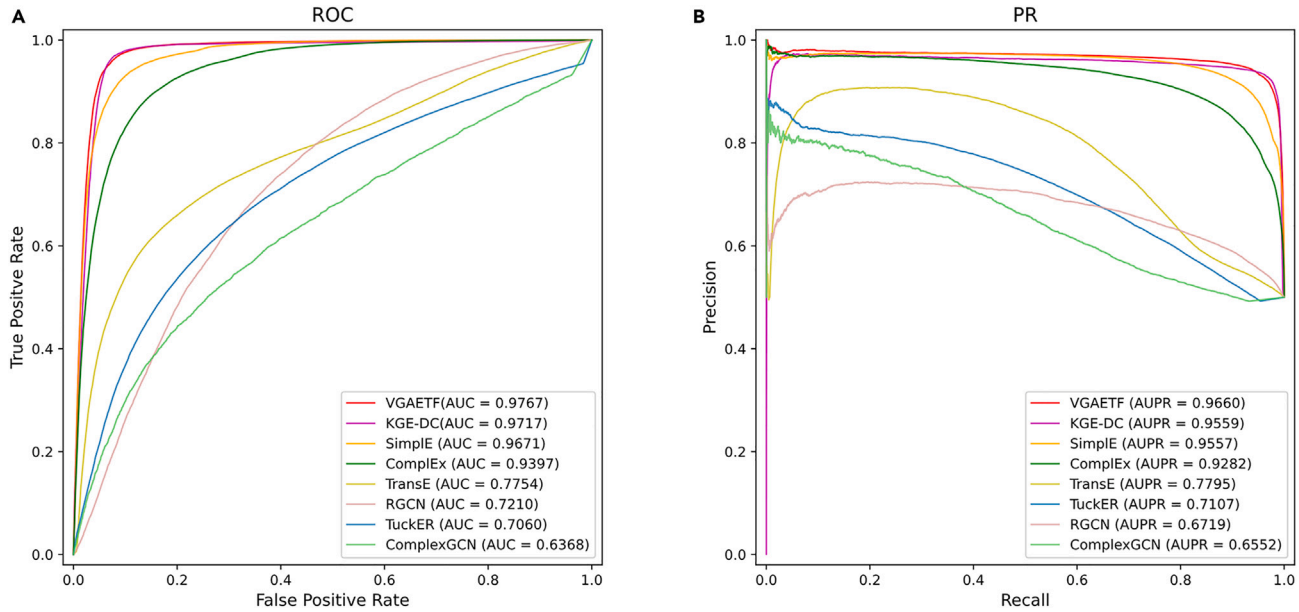


Figure 4. Performance of VGAETF as compared with other models

(A) ROC curve and AUC value of the VGAETF model and comparison methods.

(B) Precision/recall curve and AUPRC value of the VGAETF model and comparison methods.

Table 4 lists the top 10 predicted candidates for trastuzumab-containing drug combinations for breast cancer treatment. Nine of these candidates are supported by clinical studies, including paclitaxel, pertuzumab, capecitabine, docetaxel, gemcitabine, cyclophosphamide, doxorubicin, fulvestrant, and olaparib. The combination of trastuzumab and paclitaxel has been shown to significantly improve survival rates in patients with human epidermal growth factor receptor 2 (HER2)-positive breast cancer.[60] The trastuzumab-pertuzumab combination is safe and effective in the treatment of HER2-positive early-stage breast cancer.[61] The combination therapy of trastuzumab and capecitabine can produce a synergistic effect in treating breast cancer, resulting in more effective killing of cancer cells.[62] The combination of trastuzumab and docetaxel has been shown to be effective, improving survival and progression-free survival in patients with HER2-positive breast cancer.[63] The trastuzumab-gemcitabine combination shows additive or synergistic antitumor effects in human breast cancer cell lines that overexpress HER2.[64] The combination therapy of trastuzumab and cyclophosphamide is effective as adjuvant therapy in patients with stage I–III breast cancer.[65] The combination of doxorubicin and trastuzumab is used to treat early-stage HER2-positive breast cancer, and doxorubicin is a chemotherapy drug that can be used after surgery as adjuvant treatment for breast cancers.[66] Trastuzumab-fulvestrant demonstrates

Table 3. Top 10 predicted hydrochlorothiazide-containing drug combinations for treating hypertension

Rank	Drug Combination (triplets)	Evidence Source	Evidence type
1	Hydrochlorothiazide-hypertension-Spironolactone	PMID: 363334	Clinical Trial
2	Hydrochlorothiazide-hypertension-Lisinopril	PMID: 27703294	Randomized Controlled Trial
3	Hydrochlorothiazide-hypertension-Labetalol	PMID: 2860990	Clinical Trial
4	Hydrochlorothiazide-hypertension-Aliskiren	PMID: 21122057	Randomized Controlled Trial
5	Hydrochlorothiazide-hypertension-Carvedilol	PMID: 1974505	Clinical Trial
6	Hydrochlorothiazide-hypertension-Valsartan	PMID: 11800062	Clinical Trial
7	Hydrochlorothiazide-hypertension-Eplerenone	PMID: 27651039	Randomized Controlled Trial
8	Hydrochlorothiazide-hypertension-Raltegravir	N/A	N/A
9	Hydrochlorothiazide-hypertension-Amlodipine	PMID: 19470877	Randomized Controlled Trial
10	Hydrochlorothiazide-hypertension-Colchicine	N/A	N/A

Table 4. Top 10 predicted trastuzumab-containing drug combinations for treating breast cancer

Rank	Drug Combination (triplets)	Evidence Source	Evidence type
1	Trastuzumab-breast cancer-Paclitaxel	PMID: 25564897	Clinical Trial
2	Trastuzumab-breast cancer-Pertuzumab	PMID: 28581356	Clinical Trial
3	Trastuzumab-breast cancer-Capecitabine	PMID: 31825569	Randomized Controlled Trial
4	Trastuzumab-breast cancer-Docetaxel	PMID: 32171426	Clinical Trial
5	Trastuzumab-breast cancer-Gemcitabine	PMID: 12722022	Clinical Trial
6	Trastuzumab-breast cancer-Cyclophosphamide	PMID: 16978400	Clinical Trial
7	Trastuzumab-breast cancer-Cisplatin	N/A	N/A
8	Trastuzumab-breast cancer- Doxorubicin	PMID: 30621698	Clinical Trial
9	Trastuzumab-breast cancer- Fulvestrant	PMID: 34453206	Clinical Trial
10	Trastuzumab-breast cancer-Olaparib	NCT03931551	Clinical trial

effectiveness in patients with hormone receptor and HER2-positive metastatic breast cancer, and the combination therapy exhibits moderate clinical efficacy and does not exhibit significant toxicity following standard anti-HER2 treatment.[67] Trastuzumab-olaparib can be combined to treat HER2-positive breast cancer, primarily by inhibiting the growth and spread of cancer cells through the activity of HER2 and poly (ADP-ribose) polymerase (PARP) proteins.[68]

In this study, we propose a novel computational framework, VGAETF, for predicting disease-associated drug combinations. The VGAETF follows the encoder-decoder framework and employs multi-task learning, where the main task is to predict disease-related drug combinations, while the auxiliary task involves learning the representation matrix of disease-disease and drug-disease relationships. Experimental results show that VGAETF outperforms baseline methods, and case studies focusing on the prediction of novel drug pairs for the treatment of hypertension containing hydrochlorothiazide and new drug pairs for breast cancer treatment containing trastuzumab demonstrate the effectiveness of our model. In our model, we construct a drug-disease-drug multi-relational graph. Multi-relational graphs enable easy modeling of complex relationships between entities in biological systems by representing drugs and diseases as nodes and their associations as edges. Moreover, we show that the multi-task learning of our model helps improve the predictive ability of the model by studying the ablation experiments of three variants of VGAETF. In summary, the proposed VGAETF framework provides a promising approach for predicting disease-related drug combinations, and its effectiveness can be further improved with the incorporation of additional information sources.

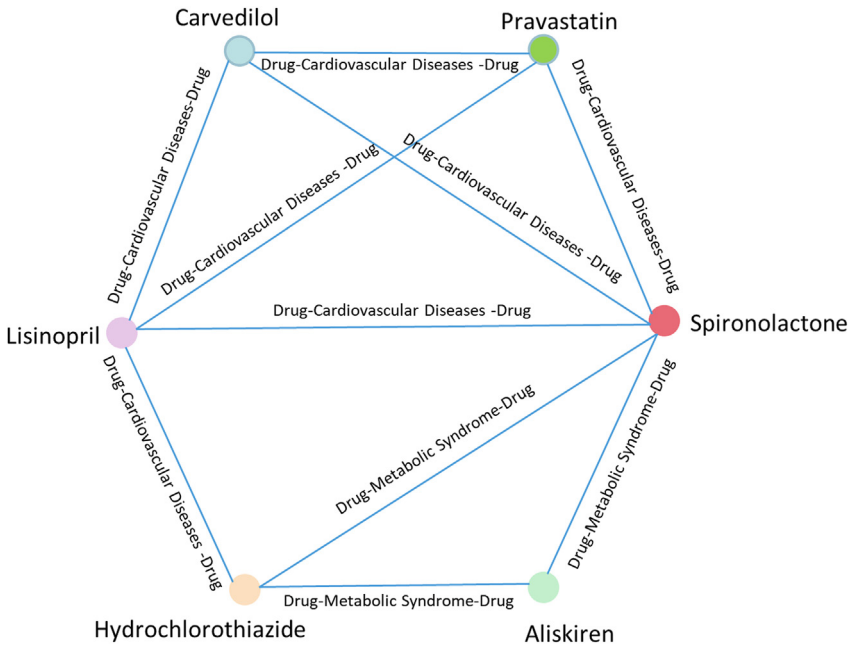


Figure 5. Visualization of the subgraph returned by GNNExplainer for explaining the spironolactone central node.

Algorithm 1. The complete procedure of VGAETF

Input: drug-disease-drug multi-relational graph $G = (\mathcal{V}, \mathcal{E}, \mathcal{R})$; drug-disease relationship matrix DR; disease-disease relationship matrix RR; train triplets set $\mathcal{T}_{train}$; test triplets set $\mathcal{T}_{test}$; maximum training epochs T;

Output: triplets prediction probability $f(s, r, o)$ in $\mathcal{T}_{test}$.

1. Initialize trainable parameters $\theta = \{X, W_0, W_1, M_r\}$.
2. t ← 1;
3. repeat
4. Learn drug embedding features Z_D with Equation 5;
5. Calculate the triplets prediction probability $f(s, r, o)$ with Equation 6;
6. Reconstruct the representation matrix of drug-disease DR and disease-disease RR with Equation 7;
7. Update drug features Z_D, disease features Z_R and the parameters of VGAETF by optimizing Equation 11;
8. util t > T or Equation 11 is converged;
9. return the triplets prediction probability $f(s, r, o)$ in $\mathcal{T}_{test}$ with Equation 6.

Limitations of the study

While VGAETF demonstrates remarkable prediction performance and interpretability, there are certain limitations to consider. Future research endeavors can concentrate on refining negative sampling methods to further improve prediction accuracy. Additionally, exploring transfer learning techniques by pre-training models on extensive and diverse knowledge bases could be beneficial. This would allow fine-tuning of the models specifically for prediction tasks. Moreover, our plans involve incorporating drug metadata into our model to enhance the learning process.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Preliminaries
  - Datasets
  - Description of VGAETF
  - Optimization
  - Experiments for performance evaluation

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108020.

ACKNOWLEDGMENTS

This work has been supported by the National Natural Science Foundation of China (No. 62002154), the Hunan Provincial Natural Science Foundation of China (No. 2021JJ40467), Research Foundation of Hunan Educational Committee (No. 20C1579), Scientific Research Startup Foundation of University of South China (No. 190XQD096), and Postgraduate Scientific Research Innovation Project of Hunan Province (CX20230968).

AUTHOR CONTRIBUTIONS

Wenyu Shan and Pingjian Ding designed the study, curated the data, and wrote the manuscript. Wenyu Shan, Cong Shen, and Pingjian Ding analyzed the results. Cong Shen, Lingyun Luo, and Pingjian Ding supervised the research. All authors discussed the results and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 18, 2023
Revised: August 26, 2023
Accepted: September 19, 2023
Published: September 22, 2023

REFERENCES

1. Lee, J.J., Lin, H.Y., Liu, D.D., and Kong, M. (2010). Emax model and interaction index for assessing drug interaction in combination studies. Front. Biosci. *2*, 582–601.
2. Nelson, H.S. (2001). Advair: combination treatment with fluticasone propionate/salmeterol in the treatment of asthma. J. Allergy Clin. Immunol. *107*, 397–416.
3. Glass, G. (2004). Cardiovascular combinations. Nat. Rev. Drug Discov. *3*, 731–732.
4. Borisy, A.A., Elliott, P.J., Hurst, N.W., Lee, M.S., Lehár, J., Price, E.R., Serbedzija, G., Zimmermann, G.R., Foley, M.A., Stockwell, B.R., and Keith, C.T. (2003). Systematic discovery of multicomponent therapeutics. Proc. Natl. Acad. Sci. USA *100*, 7977–7982.
5. Sun, M., Zhao, S., Gilvary, C., Elemento, O., Zhou, J., and Wang, F. (2020). Graph convolutional networks for computational drug development and discovery. Brief. Bioinform. *21*, 919–935.
6. Dang, C.V., Reddy, E.P., Shokat, K.M., and Soucek, L. (2017). Drugging the'undruggable'cancer targets. Nat. Rev. Cancer *17*, 502–508.
7. Jiang, P., Huang, S., Fu, Z., Sun, Z., Lakowski, T.M., and Hu, P. (2020). Deep graph embedding for prioritizing synergistic anticancer drug combinations. Comput. Struct. Biotechnol. J. *18*, 427–438.
8. Cheng, F., Kovács, I.A., and Barabási, A.L. (2019). Network-based prediction of drug combinations. Nat. Commun. *10*, 1197.
9. Li, T., Shetty, S., Kamath, A., Jaiswal, A., Jiang, X., Ding, Y., and Kim, Y. (2023). CancerGPT: Few-shot Drug Pair Synergy Prediction using Large Pre-trained Language Models. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.10946.
10. Zhang, P., and Tu, S. (2022). A Knowledge Graph Embedding-Based Method for Predicting the Synergistic Effects of Drug Combinations (IEEE), pp. 1974–1981.
11. Wu, L., Gao, J., Zhang, Y., Sui, B., Wen, Y., Wu, Q., Liu, K., He, S., and Bo, X. (2023). A hybrid deep forest-based method for predicting synergistic drug combinations. Cell Rep. Methods *3*, 100411.
12. Wang, H., Lee, D.-K., Chen, K.-Y., Chen, J.-Y., Zhang, K., Silva, A., Ho, C.-M., and Ho, D. (2015). Mechanism-independent optimization of combinatorial nanodiamond and unmodified drug delivery using a phenotypically driven platform technology. ACS Nano *9*, 3332–3344.
13. Gao, Z., Ding, P., and Xu, R. (2022). Kg-predict: a knowledge graph computational framework for drug repurposing. J. Biomed. Inform. *132*, 104133.
14. Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P.S. (2022). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Trans. Neural Netw. Learn. Syst. *33*, 494–514.
15. Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G.d., Gutierrez, C., Kirrane, S., Gayo, J.E.L., Navigli, R., Neumaier, S., et al. (2021). Knowledge graphs. ACM Comput. Surv. *54*, 1–37.
16. Mohamed, S.K., Nounu, A., and Nováček, V. (2021). Biological applications of knowledge graph embedding models. Brief. Bioinform. *22*, 1679–1693.
17. Nováček, V., and Mohamed, S.K. (2020). Predicting polypharmacy side-effects using knowledge graph embeddings. AMIA Jt. Summits Transl. Sci. Proc. *2020*, 449–458.
18. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. Adv. Neural Inf. Process. Syst. *26*.
19. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. (2016). Complex Embeddings for Simple Link Prediction (PMLR), pp. 2071–2080.
20. Lacroix, T., Usunier, N., and Obozinski, G. (2018). Canonical Tensor Decomposition for Knowledge Base Completion (PMLR), pp. 2863–2872.
21. Balažević, I., Allen, C., and Hospedales, T.M. (2019). Tucker: Tensor factorization for knowledge graph completion. Preprint at arXiv. https://doi.org/10.18653/v1/D19-1522.
22. Zhang, Z., Cai, J., Zhang, Y., and Wang, J. (2020). Learning hierarchy-aware knowledge graph embeddings for link prediction. Preprint at arXiv *34*, 3065–3072.
23. Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., and Welling, M. (2018). Modeling Relational Data with Graph Convolutional Networks (Springer), pp. 593–607.
24. Zeb, A., Saif, S., Chen, J., Haq, A.U., Gong, Z., and Zhang, D. (2022). Complex graph convolutional network for link prediction in knowledge graphs. Expert Syst. Appl. *200*, 116796.
25. Kipf, T.N., and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1609.02907.
26. Shtar, G., Azulay, L., Nizri, O., Rokach, L., and Shapira, B. (2022). CDCDB: A large and continuously updated drug combination database. Sci. Data *9*, 263.
27. Ioannidis, V., Song, X., Manchanda, S., Li, M., Pan, X., Zheng, D., Ning, X., Zeng, X., and Karypis, G. (2020). DRKG-Drug Repurposing Knowledge Graph for COVID-19.
28. Kingma, D.P., and Welling, M. (2014). Stochastic Gradient VB and the Variational Auto-Encoder, p. 121.
29. Kipf, T.N., and Welling, M. (2016). Variational graph auto-encoders. Preprint at arXiv. https://doi.org/10.48550/arXiv.1611.07308.
30. Bowers, A.J., and Zhou, X. (2019). Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. J. Educ. Stud. Placed A. T. Risk *24*, 20–46.
31. Qi, Q., Luo, Y., Xu, Z., Ji, S., and Yang, T. (2021). Stochastic optimization of areas under precision-recall curves with provable convergence. Adv. Neural Inf. Process. Syst. *34*, 1752–1765.
32. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic Differentiation in Pytorch.
33. Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., Zhou, J., Ma, C., Yu, L., and Gai, Y. (2019). Deep graph library: A graph-centric, highly-performant package for graph neural networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.1909.01315.
34. Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., Catasta, M., and Leskovec, J. (2020). Open graph benchmark: Datasets for machine learning on graphs. Adv. Neural Inf. Process. Syst. *33*, 22118–22133.
35. Glorot, X., and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In JMLR Workshop and Conference Proceedings, pp. 249–256.
36. Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.
37. Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.
38. Asami, K., Ando, M., Nishimura, T., Yokoi, T., Tamura, A., Minato, K., Mori, M., Ogushi, F., Yamamoto, A., Yoshioka, H., et al. (2022). A randomized phase II study of docetaxel or pemetrexed with or without the continuation of gefitinib after disease progression in elderly patients with non-small cell lung cancer harboring EGFR mutations (JMTO LC12-01). Thorac. Cancer *13*, 1827–1836.
39. Hanna, N., Shepherd, F.A., Fossella, F.V., Pereira, J.R., De Marinis, F., Von Pawel, J., Gatzemeier, U., Tsao, T.C.Y., Pless, M., Muller, T., et al. (2004). Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. J. Clin. Oncol. *22*, 1589–1597.
40. Shiraishi, Y., Daga, H., Ikeda, S., Hata, A., Mizutani, H., Sakamoto, T., Saito, H., Hataji, O., Tanaka, H., and Horiike, A. (2019). A multicenter, open label, randomized phase III study of atezolizumab with platinum-pemetrexed and with or without bevacizumab for patients with advanced nonsquamous non-small cell lung cancer (WJOG11218L APPLE Study). Am. Soc. Clin. Oncol. *37*, TPS9125.
41. Lam, T.C., Tsang, K.C., Choi, H.C., Lee, V.H., Lam, K.O., Chiang, C.L., So, T.H., Chan, W.W., Nyaw, S.F., Lim, F., et al. (2021). Combination atezolizumab, bevacizumab, pemetrexed and carboplatin for metastatic EGFR mutated NSCLC after TKI failure. Lung Cancer *159*, 18–26.
42. Kazemi, S.M., and Poole, D. (2018). Simple embedding for link prediction in knowledge graphs. Adv. Neural Inf. Process. Syst. *31*.
43. Hitchcock, F.L. (1927). The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. *6*, 164–189.

44. Ernst, M.E., and Fravel, M.A. (2022). Thiazide and the thiazide-like diuretics: review of hydrochlorothiazide, chlorthalidone, and indapamide. Am. J. Hypertens. *35*, 573–586.
45. Schrijver, G., and Weinberger, M.H. (1979). Hydrochlorothiazide and spironolactone in hypertension. Clin. Pharmacol. Ther. *25*, 33–42.
46. Sukalo, A., Deljo, D., Krupalija, A., Zjajo, N., Kos, S., Curic, A., Divkovic, G., Hubjar, S., Smailagic, M., Hodzic, E., et al. (2016). Treatment of hypertension with combination of lisinopril/hydrochlorothiazide. Med. Arch. *70*, 299–302.
47. Miller, E. (1991). Introduction to lisinopril-hydrochlorothiazide combination. J. Hum. Hypertens. *5*, 49–51.
48. Group, L.H.M.S. (1985). Labetalol and hydrochlorothiazide in hypertension. Clin. Pharmacol. Ther. *38*, 24–27.
49. Chrysant, S.G. (2008). Aliskiren–hydrochlorothiazide combination for the treatment of hypertension. Expert Rev. Cardiovasc Ther. *6*, 305–314.
50. Schmieder, R.E., Philipp, T., Gueredіaga, J., Gorostidi, M., Bush, C., and Keefe, D.L. (2009). Aliskiren-based therapy lowers blood pressure more effectively than hydrochlorothiazide-based therapy in obese patients with hypertension: sub-analysis of a 52-week, randomized, double-blind trial. J. Hypertens. *27*, 1493–1501.
51. Widmann, L., Van der Does, R., Hörrmann, M., and Machwirth, M. (1990). Safety and antihypertensive efficacy of carvedilol and atenolol alone and in combination with hydrochlorothiazide. Eur. J. Clin. Pharmacol. *38*, S143–S146.
52. Schmidt, A., Adam, S.A., Kolloch, R., Weidinger, G., and Handrock, R. (2001). Antihypertensive effects of valsartan/hydrochlorothiazide combination in essential hypertension. Blood Press. *10*, 230–237.
53. Sison, J., Vega, R.M.R., Dayi, H., Bader, G., and Brunel, P. (2018). Efficacy and effectiveness of valsartan/amlodipine and valsartan/amlodipine/hydrochlorothiazide in hypertension: randomized controlled versus observational studies. Curr. Med. Res. Opin. *34*, 501–515.
54. Karashima, S., Yoneda, T., Kometani, M., Ohe, M., Mori, S., Sawamura, T., Furukawa, K., Yamagishi, M., and Takeda, Y. (2016). Angiotensin II receptor blocker combined with eplerenone or hydrochlorothiazide for hypertensive patients with diabetes mellitus. Clin. Exp. Hypertens. *38*, 565–570.
55. Calhoun, D.A., Lacourcière, Y., Chiang, Y.T., and Glazer, R.D. (2009). Triple antihypertensive therapy with amlodipine, valsartan, and hydrochlorothiazide: a randomized clinical trial. Hypertension *54*, 32–39.
56. Julius, S. (1988). Amlodipine in hypertension: an overview of the clinical dossier. J. Cardiovasc. Pharmacol. *12*, S27–S33.
57. Ying, R., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. (2019). Gnnexplainer: Generating explanations for graph neural networks. Adv. Neural Inf. Process. Syst. *32*, 9240–9251.
58. Warrens, H., Banerjee, D., and Herzog, C.A. (2022). Cardiovascular Complications of Chronic Kidney Disease: An Introduction. Eur. Cardiol. *17*, e13.
59. Sinha, A.D., and Agarwal, R. (2015). The complex relationship between CKD and ambulatory blood pressure patterns. Adv. Chronic Kidney Dis. *22*, 102–107.
60. Tolaney, S.M., Barry, W.T., Dang, C.T., Yardley, D.A., Moy, B., Marcom, P.K., Albain, K.S., Rugo, H.S., Ellis, M., Shapira, I., et al. (2015). Adjuvant paclitaxel and trastuzumab for node-negative, HER2-positive breast cancer. N. Engl. J. Med. *372*, 134–141.
61. Von Minckwitz, G., Procter, M., De Azambuja, E., Zardavas, D., Benyunes, M., Viale, G., Suter, T., Arahmani, A., Rouchet, N., Clark, E., et al. (2017). Adjuvant pertuzumab and trastuzumab in early HER2-positive breast cancer. N. Engl. J. Med. *377*, 122–131.
62. Murthy, R.K., Loi, S., Okines, A., Paplomata, E., Hamilton, E., Hurvitz, S.A., Lin, N.U., Borges, V., Abramson, V., Anders, C., et al. (2020). Tucatinib, trastuzumab, and capecitabine for HER2-positive metastatic breast cancer. N. Engl. J. Med. *382*, 597–609.
63. Swain, S.M., Miles, D., Kim, S.-B., Im, Y.-H., Im, S.-A., Semiglazov, V., Ciruelos, E., Schneeweiss, A., Loi, S., Monturus, E., et al. (2020). Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA): end-of-study results from a double-blind, randomised, placebo-controlled, phase 3 study. Lancet Oncol. *21*, 519–530.
64. O'shaughnessy, J. (2003). Gemcitabine and Trastuzumab in Metastatic Breast Cancer, *2* (Elsevier), pp. 22–26.
65. Orlando, L., Cardillo, A., Ghisini, R., Rocca, A., Balduzzi, A., Torrisi, R., Peruzzotti, G., Goldhirsch, A., Pietri, E., and Colleoni, M. (2006). Trastuzumab in combination with metronomic cyclophosphamide and methotrexate in patients with HER-2 positive metastatic breast cancer. BMC Cancer *6*, 225–228.
66. Gavilá, J., Oliveira, M., Pascual, T., Perez-Garcia, J., Gonzàlez, X., Canes, J., Paré, L., Calvo, I., Ciruelos, E., and Muñoz, M. (2019). Safety, activity, and molecular heterogeneity following neoadjuvant non-pegylated liposomal doxorubicin, paclitaxel, trastuzumab, and pertuzumab in HER2-positive breast cancer (Opti-HER HEART): an open-label, single-group, multicenter, phase 2 trial. BMC Med. *17*, 1–12.
67. Torrisi, R., Palumbo, R., De Sanctis, R., Vici, P., Bianchi, G.V., Cortesi, L., Leonardi, V., Gueli, R., Fabi, A., Valerio, M.R., et al. (2021). Fulvestrant and trastuzumab in patients with luminal HER2-positive advanced breast cancer (ABC): an Italian real-world experience (HERMIONE 9). Breast Cancer Res. Treat. *190*, 103–109.
68. Alés-Martínez, J., Morales, S., Fernández Abad, M., Sánchez-Rovira, P., Salvador Bofill, F., Lahuerta, A., García-Sáenz, J., Garau Llinas, I., Díaz Redondo, T., Ferrer, N., et al. (2019). Effectiveness of olaparib plus trastuzumab in HER2 [+], BRCA–mutated (BRCAm) or homologous recombination deficient (HRD) advanced breast cancer (ABC) patients (pts). The OPHELIA study. Ann. Oncol. *30*, v139–v140.
69. Pourkavoos, N. (2012). Unique risks, benefits, and challenges of developing drug-drug combination products in a pharmaceutical industrial setting. Comb. Prod. Ther. *2*, 2–31.
70. Luan, S., Hua, C., Lu, Q., Zhu, J., Chang, X.-W., and Precup, D. (2022). When Do We Need GNN for Node Classification?. Preprint at arXiv. https://doi.org/10.48550/arXiv.2210.16979.
71. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proc. IEEE *86*, 2278–2324.
72. Kullback, S., and Leibler, R.A. (1951). On information and sufficiency. Ann. Math. Statist. *22*, 79–86.
73. Prokhorov, V., Shareghi, E., Li, Y., Pilehvar, M.T., and Collier, N. (2019). On the importance of the Kullback-Leibler divergence term in variational autoencoders for text generation. Preprint at arXiv. https://doi.org/10.48550/arXiv.1909.13668.
74. Chen, M., Wei, Z., Ding, B., Li, Y., Yuan, Y., Du, X., and Wen, J.-R. (2020). Scalable graph neural networks via bidirectional propagation. Adv. Neural Inf. Process. Syst. *33*, 14556–14566.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
CDCDB	Shtar, G. et al.[26]	https://doi.org/10.1038/s41597-022-01360-z
DRKG	Ioannidis, V. et al.[27]	https://github.com/gnn4dr/DRKG/
Software and algorithms		
VGAETF	this study	https://github.com/hhhwy/VGAETF
TransE	Bordes, A. et al.[18]	https://github.com/snap-stanford/ogb/tree/master
ComplEx	Trouillon, T. et al.[19]	https://github.com/snap-stanford/ogb/tree/master
SimplE	Kazemi, S. M. and D. Poole[42]	https://github.com/snap-stanford/ogb/tree/master
TuckER	Balažević, I. et al.[21]	https://github.com/ibalazevic/TuckER
R-GCN	Schlichtkrull, M. et al.[23]	https://github.com/snap-stanford/ogb/tree/master
ComplexGCN	Zeb, A. et al.[24]	https://github.com/Aazeb/ComplexGCN
KGE-DC	Zhang, P. and S. Tu[10]	https://github.com/yushenshashen/KGE-DC

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Pingjian Ding (dpj@usc.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data reported in this paper is publicly available on GitHub (https://github.com/hhhwy/VGAETF)
- All original code has been deposited at GitHub and is publicly available as of the date of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

METHOD DETAILS

Preliminaries

The development of synergistic drug combinations may largely be based on the rationale of the treatment of two closely related diseases,[69] and GNN mostly follows the assumption of learning to assign similar labels to nodes that are closely connected.[70] Therefore, in this study, we developed a GNN-based method on a multi-relational graph, where two drugs represent entities and diseases represent relationships in the triplets. We explain the disease-related drug combinations prediction task in Figure S1. To predict the disease-related drug combinations, we incorporate additional known triplets associated with the drugs and diseases to learn the embedding vectors for them. Subsequently, these learned vectors are used to generate a score for each drug-disease-drug triplet by the scoring function of tensor decomposition. As shown in Figure S1, in the context of known drug combinations, carvedilol can be combined with spironolactone or pravastatin to treat cardiomyopathies. In the context of drug indication and disease comorbidities, spironolactone and pravastatin can be used for treating high blood pressure and high cholesterol respectively, while hypertension and hyperlipidemia are all risk factors of heart disease. Taken together, spironolactone and pravastatin are intended to be used in combination to treat or prevent diseases (such as heart disease) closely related to hyperlipidemia hypertension, and cardiomyopathies.

Datasets

We obtained a dataset from CDCDB,[26] a comprehensive drug combination database containing more than 40,795 drug combinations. We extracted the DrugBank identifiers (IDs) that are related to the FDA Orange Book and the Medical Subject Headings (MeSH) terms that describe the conditions that are treated in the study from our database. The data schema ($drug_i$, $disease_k$, $drug_j$) shows that combining $drug_i$ and $drug_j$ can treat $disease_k$. Among the labeled "Efficacious" drug pairs for one or more diseases, there are a total of 1,105 individual drugs, 1,283 diseases, and 4,709 pairs of drugs. We constructed 59,659 positive samples of drug-disease-drug triplets using this information. In this study, we randomly selected 90% of the positive samples as training data, while reserving the remaining 10% as testing data. Based on the drugs and diseases obtained from the training data, we obtained 26,425 drug-disease associations in the association data. We construct a binary matrix to denote known drug-disease associations that values of 1 and 0 indicate verified associations and unverified associations between drugs and targets, respectively. We collate 259 disease-disease associations from DRKG.[27] This dataset contains disease-resembles-disease edges that are based on the co-occurrence of disease terms in PubMed. We also construct a binary matrix that includes two values 1 and 0, denoting validated and unvalidated associations between diseases and diseases, respectively. More details can be found in Table 1.

Description of VGAETF

VGAETF encoder

We follow the encoder approach like VGAE[29] to learn the feature representation of the node. We take drug combinations of the drug-disease-drug triplets as the input data for the VGAE encoder. Finally, we can get the low-dimensional representations of drugs as $Z_D \in \mathbb{R}^{N\times F}$ where F is the dimension of the final output of the encoder. The drug input features are $X \in \mathbb{R}^{N\times N}$ where N is the total number of drugs. The VAE model consists of two parts: encoding and decoding.

The VGAE framework leverages the variational auto-encoder to learn graph-structured data by naturally incorporating node features using GCN. This enables the model to generate interpretable latent representations of undirected graphs based on data distribution. The encoder takes the adjacency matrix A (every node is connected to itself) and the feature matrix X as inputs and uses two GCN layers to obtain a latent variable z. The two layers of GCN in VGAE are defined as follows:

$$\mathrm{GCN}(\mathrm{X},\mathrm{A}) = \tilde{\mathrm{A}}\mathrm{ReLU}(\tilde{\mathrm{A}}\mathrm{X}\mathrm{W}_0)\mathrm{W}_1. \quad \text{(Equation 1)}$$

where W_0 and W_1 are the trainable weight matrices of the first and second layers. $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the symmetrically normalized new adjacency matrix with A added self-connections, and D is the degree matrix of A. W_0 is shared by GCN_σ and GCN_μ. Specifically, VGAE employs two graph neural networks to fit the μ and σ vectors for each node i:

$$\mu = \mathrm{GCN}_\mu(\mathrm{X},\mathrm{A}). \quad \text{(Equation 2)}$$

$$log\,\sigma = \mathrm{GCN}_\sigma(\mathrm{X},\mathrm{A}). \quad \text{(Equation 3)}$$

The latent variable z is then computed as follows:

$$z = \mu + \varepsilon \cdot \sigma. \quad \text{(Equation 4)}$$

where ε follows $N(0,1)$. The encoder can also be written as follows:

$$q(\mathrm{z}_i|\,\mathrm{X},\mathrm{A}) = N\left(z_i|\,\mu_i, \mathrm{diag}\left(\sigma_i^2\right)\right). \quad \text{(Equation 5)}$$

VGAETF decoder

We follow the same tensor factorization method as R-GCN[23] to calculate the prediction scores of triplet facts using the final drug representations Z_D learned from the VGAE encoder and a trainable diagonal matrix M_k associated with diseases. Also, we use the inner product as a decoder to learn the representation matrix of disease comorbidities and drug-disease relationships using the node representations of drugs and diseases.

1. **tensor factorization decoder:** For the tensor factorization decoder, f assigns to each triplet (s, r, o) a score that represents the valid probability of a triplet:

$$f(s,r,o) = z_s^T M_r z_o. \quad \text{(Equation 6)}$$

where the trainable diagonal matrix M_r is specific to the disease type r, while s and o refer to the drug entity in the triplet.

2. **inner product decoder:** The decoder computes the reconstructed adjacency matrix by computing the inner product between latent variables z_i and z_j of the nodes, and applying the logistic sigmoid function $\sigma(\cdot)$ as follows:

$$\varphi\left(z_i, z_j\right) = \sigma\left(z_i z_j^T\right). \quad \text{(Equation 7)}$$

Optimization

To account for the lack of negative samples in the known drug-disease-drug triplets, our model generates negative samples during training. We achieve this by randomly corrupting the subject or object of each positive example. The resulting set of negative samples is combined with the real triplets to form the total training set $\mathcal{T}$. We then use cross-entropy loss function[71] to train the model, trying to give a higher probability to seen edges and a lower probability to unseen edges. Our loss function is designed as follows:

$$\mathcal{L}_1 = -\frac{1}{|S|}\sum_{(s,r,o,y)\in\mathcal{T}} y\log l(f(s,r,o)) + (1-y)\log(1-l(f(s,r,o))). \quad \text{(Equation 8)}$$

where $|S|$ is the number of training triplets. The score $f(s,r,o)$ is normalized between 0 and 1 using the logistic sigmoid function $l(x) = 1/(1+\exp(-x))$. The indicator variable y is set to 1 for positive triplets and 0 for negative ones.

To predict drug combinations using auxiliary information, we optimize our model using a combination of loss functions. Our optimization function, denoted as $\mathcal{L}$, includes the loss function $\mathcal{L}_1$, $\mathcal{L}_2$ for disease-disease relationships, and $\mathcal{L}_3$ for drug-disease relationships. Additionally, $\mathcal{L}$ incorporates the KL divergence[72] between $q(Z|X,\mathcal{E})$ and $p(Z)$. KL divergence serves as a regularization term that encourages the learned latent space to follow a predefined prior distribution and prevents overfitting to the training data.[73] Formally, $\mathcal{L}_2$ and $\mathcal{L}_3$ can be defined as:

$$\mathcal{L}_2 = -\frac{1}{|RR|}\sum_{(r_1,r_2)\in RR}\log\left(\psi\left(R_{r_1},R_{r_2}\right)\right) \quad \text{(Equation 9)}$$

$$\mathcal{L}_3 = -\frac{1}{|DR|}\sum_{(d,r)\in DR}\log(\psi(D_d,R_r)) \quad \text{(Equation 10)}$$

where RR stands for all disease pairs in the disease-disease relationship and DR stands for all drug-disease pairs in the drug-disease relationship.

We integrate $\mathcal{L}$ to train the model, i.e.,

$$\mathcal{L} = \mathcal{L}_1 + (1-\alpha)\cdot\mathcal{L}_2 + \alpha\cdot\mathcal{L}_3 - KL\left[q(Z|X,\mathcal{E})\|p(Z)\right] \quad \text{(Equation 11)}$$

where α is a hyperparameter that trades off the two loss functions $\mathcal{L}_2$ and $\mathcal{L}_3$ in the auxiliary task. Our model learns embeddings for drugs and diseases using the Adam optimizer[36] to minimize the loss function.

Similar to the calculation process in Chen et al.,[74] we analyze the computational complexity for two key components of our model. For the training and inference time complexity of the two-layer GCN can be bounded by $O(|E|N + |E|F + N^3 + N^2F)$, where $|E|$ represents the number of edges of the undirected graph. For the decoder, the time complexity is $O(|\mathcal{T}|F + 2F)$. A complete description of the procedure of VGAETF is presented in Algorithm 1.

Experiments for performance evaluation

We evaluated the performance of the VGAETF model using the area under the receiver operating characteristic (ROC) curve (AUC)[30] and the area under the precision/recall curve (AUPRC).[31] We implemented the VGAETF models with PyTorch[32] and DGL packages.[33] The publicly available source codes were used to implement all baseline methods. The best or default parameters of each method are used. The implementation of TransE, ComplEx, SimplE, and RGCN models are provided by the Open Graph Benchmark.[34] We use the model implementation of TuckER, KGE-DC, and ComplexGCN provided by the authors. Details are in the key resources table.

In this study, we utilized grid search to determine the optimal parameters for our model. Specifically, we initialized the disease embeddings using Xavier initialization[35] and optimized the training loss using the Adam solver.[36] Various hyperparameters were taken into account in our model. The hyperparameters in the VGAETF model, which include the learning rate, the number of units in hidden layer 1 and hidden layer 2, and the balance factor α, were considered. The hyper-parameter α balances the contributions of the representation matrices to learn disease-disease relationships and drug-disease relationships in the auxiliary task. To prevent overfitting and improve efficiency, early stopping was implemented during training, and the number of training epochs was fixed at 1200. The optimal hyperparameters were selected and highlighted in bold in Table 2.