JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

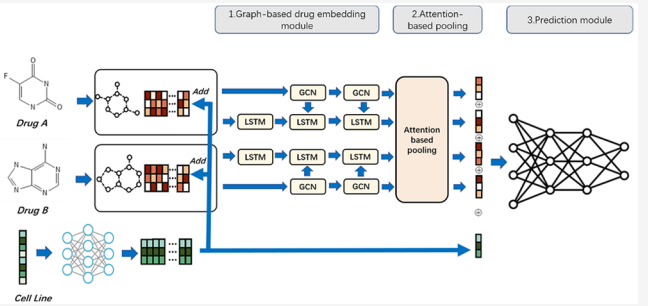
AttenSyn: An Attention-Based Deep Graph Neural Network for Anticancer Synergistic Drug Combination Prediction

Tianshuo Wang, Ruheng Wang, and Leyi Wei*

Cite This: J. Chem. Inf. Model. 2024, 64, 2854−2862 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Identifying synergistic drug combinations is fundamentally important to treat a variety of complex diseases while avoiding severe adverse drug−drug interactions. Although several computational methods have been proposed, they highly rely on handcrafted feature engineering and cannot learn better interactive information between drug pairs, easily resulting in relatively low performance. Recently, deep-learning methods, especially graph neural networks, have been widely developed in this area and demonstrated their ability to address complex biological problems. In this study, we proposed AttenSyn, an attention-based deep graph neural network for accurately predicting synergistic drug combinations. In particular, we adopted a graph neural network module to extract high-latent features based on the molecular graphs only and exploited the attention-based pooling module to learn interactive information between drug pairs to strengthen the representations of drug pairs. Comparative results on the benchmark datasets demonstrated that our AttenSyn performs better than the state-of-the-art methods in the prediction of anticancer synergistic drug combinations. Additionally, to provide good interpretability of our model, we explored and visualized some crucial substructures in drugs through attention mechanisms. Furthermore, we also verified the effectiveness of our proposed AttenSyn on two cell lines by visualizing the features of drug combinations learnt from our model, exhibiting satisfactory generalization ability.



INTRODUCTION

In recent decades, the number of available anticancer drugs has been increasing rapidly, accompanied by an evolving understanding of the biological complexity of malignant tumors. However, in cancer, multiple cellular mechanisms are often altered in the cell; therefore, treating them with a single drug and focusing on a single target is usually ineffective. Compared to the traditional treatment mode of "single disease, single drug, and single target", combination therapy has the potential to increase treatment efficacy, reduce host toxicity and adverse side effects, and overcome drug resistance. Drug combinations are widely used to treat a variety of complex diseases, such as hypertension,[1] infectious diseases,[2] and cancer.[3] However, some antagonistic effects and even severe adverse drug−drug interactions occur when using some drug combinations, which not only are ineffective at enhancing the curative effect but also threaten the patient's health. Therefore, it is crucial to accurately discover synergistic drug combinations for specific diseases.

Traditional experimental methods to screen the combinations of synergistic antitumor drugs are very challenging in terms of time, efficiency, and cost, which are far from meeting the urgent need for anticancer drugs. Even high-throughput screens are not feasible due to the vast number of drug combinations. Thanks to the rapid development of related databases and machine-learning technologies, researchers have made great breakthroughs in drug-related interaction prediction.[4−6] Also, various computational methods have been proposed for drug combination prediction in the past decade, which greatly reduces the screening space of drug combinations. For example, in the streptozocin-induced neuropathic pain model in mice, the maximal antiallodynic effect of a new derivative of dihydrofuran-2-one (LPP1) used in combination with pregabalin has been successfully predicted by using the support vector machine (SVM) and random forest (RF) algorithms.[7,8] Moreover, Liu et al.[9] used the features extracted from the random walk algorithm with a restart on the drug−protein heterogeneous network and a gradient tree boosting classifier to predict new drug combinations. Later, Pivetta et al.[10] employed an artificial neural-network-based model to predict the synergism of anticancer drugs. In addition, Zhang and Yan[11] considered pharmacological data and applied field-aware factorization machines to analyze and predict potential synergistic drug combinations. Apart from these methods,

Special Issue: Machine Learning in Bio-cheminformatics

Received: May 10, 2023
Published: August 11, 2023

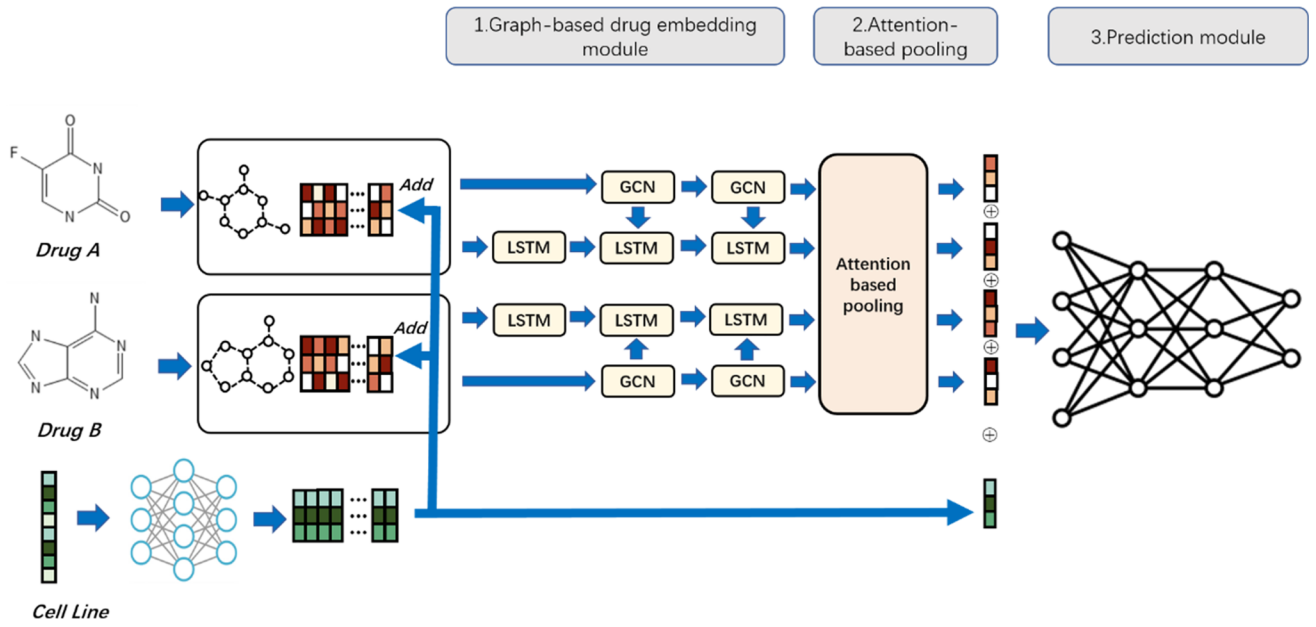


Figure 1. The overall architecture of our proposed method. Given a pair of drugs and the corresponding cell-line type, we first transform the SMILES profiles of the two drugs into molecular graphs with initial node features and graph structures. Also, we add the cell-line features obtained from CCLE into the feature of each node in the drug molecular graphs. Then we employ the graph-based drug-embedding module with several GCN and LSTM layers to learn better representations of each node and then use the attention-based pooling module to obtain the graph-level features. Finally, we exploit the prediction module to integrate the representations of two drugs with features of the cell line and make the prediction of synergistic drug combinations.

Janizek et al.[12] proposed TreeCombo, a machine-learning-based model using the XGBoost algorithm to predict the synergistic score of drug pairs. Chen et al. developed NLLSS, a novel algorithm termed "Network-based Laplacian regularized Least Square Synergistic" (NLLSS) drug combination prediction to predict potential antifungal synergistic drug combinations by integrating different kinds of information.[13] The above methods have boosted the use of machine-learning solutions for drug combination prediction. However, these traditional machine-learning workflows have some drawbacks. For instance, training a good model normally requires strong professional knowledge of handcrafted feature engineering and machine-learning algorithms, limiting their usability in real applications to some extent. Moreover, they cannot support fast large-scale prediction due to their relatively low computational ability.

Recently, with the rapid development of deep learning and the release of large-scale drug combination datasets, it has been possible to predict drug combinations using deep-learning methods. For example, Preuer et al.[14] proposed DeepSynergy, which combined the chemical descriptors of drugs and gene expression of cell lines to predict the potential drug synergies. Liu and Xie[15] proposed TranSynergy, considered the network information such as gene−gene interaction networks and drug−target associations, and applied the transformer architecture to the synergistic prediction of drugs. Moreover, Kuru et al.[16] trained two parallel subnetworks to learn drug-specific representation on a particular cell line and make predictions. Later, Su et al.[17] proposed SRDFM, integrating a factorization machine component with a deep neural network for single-drug and synergistic drug combination recommendation. Furthermore, Lin et al.[18] proposed EC-DFR, which uses physicochemical properties, molecular fingerprints, cell-line-specific drug-induced gene expression profiles as features, and a cascade-based deep forest model for drug combination prediction. Li et al. proposed SNRMPACDC, using siamese network and random matrix projection for anticancer drug combination prediction.[19] These methods make the use of deep learning more convenient for synergistic drug combination prediction to some extent. However, there remain some drawbacks that need to be addressed. First, most of the existing methods cannot meet the high demand of the research community since they possess relatively lower performances. Second, few of them have an interpretation of their models and detect what their models learn during the training process.

In recent years, graph neural networks (GNN) have achieved remarkable success in many real-world applications such as drug discovery. Many studies have used graph neural networks on the molecular graph to extract molecular features for drug combination prediction. For example, Wang et al.[20] proposed a GNN-based deep-learning network called DeepDDS, applying a graph convolutional network (GCN) and graph attention network (GAT) to extract the drug-embedding vectors for identifying drug combinations. Jin et al.[21] proposed ComboNet, which employed a directional message-passing neural network (DMPNN) to learn a continuous representation of a molecule for identifying synergistic drug combinations for treating COVID-19. Additionally, Hu et al.[22] developed a deep graph neural network model named DTSyn and used multihead attention mechanism to identify novel drug combinations. Although graph neural networks have demonstrated their ability to solve drug synergy prediction and achieved success to some extent in terms of performance improvement, there still exist some limitations. First, although some adverse drug−drug interaction methods considered the interactive information between drug pairs,[23,24] existing synergistic drug combination prediction methods focus on the information extraction of a single drug but ignore the importance of the interactive information between drug pairs. Second, most of them treat

https://doi.org/10.1021/acs.jcim.3c00709
J. Chem. Inf. Model. 2024, 64, 2854−2862

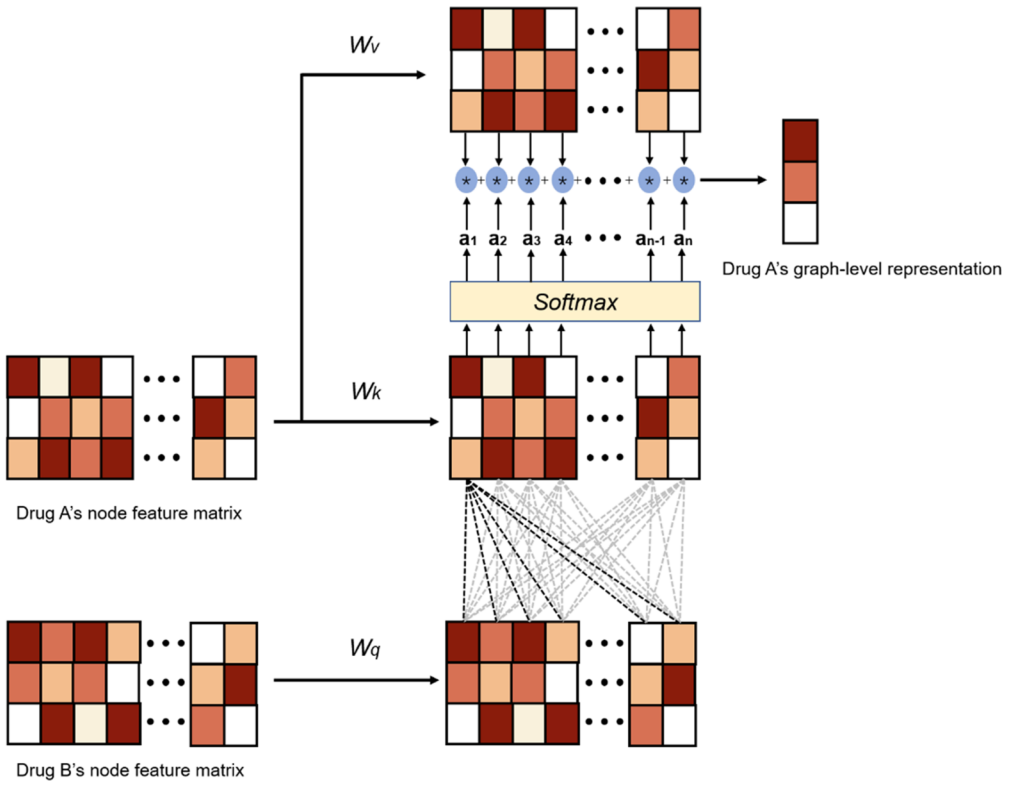


Figure 2. The overall computational steps of attention-based pooling. The graph-level representation is calculated by the weighted sum of all nodes' embeddings according to the attention scores.

each substructure in a drug molecule as equally important and cannot detect important ones for the prediction of synergistic drug combinations.

To tackle the above problems, in this study, we proposed an attention-based deep graph neural network named AttenSyn to predict the synergistic effect of drug combinations. Specifically, our proposed model incorporates several novel features as follows. First, we employ a deep graph neural network to learn and extract high-latent features automatically rather than using manual feature profiles from handcrafted feature engineering. Second, with the attention-based pooling module, we can not only learn the interactive information between drug pairs but also detect the important chemical substructures in drugs for the identification of anticancer synergistic drug combinations. Third, comparative results with the existing methods on the benchmark dataset show that our proposed model outperforms not only classical machine-learning methods but also deep-learning methods, demonstrating that our AttenSyn has great potential to be a powerful and practically useful deep-learning tool for anticancer synergistic drug combinations prediction.

METHODS AND MATERIALS

Dataset. To compare the performance of our proposed model with the state-of-the-art methods, we collected the drug combination dataset constructed by O'Neil et al. as our benchmark dataset.[25] The dataset contains 23,052 triplets, where each triplet comprises two drugs and a cancer cell line.[22] There are 39 cancer cell lines and 38 unique drugs in the dataset, and these drugs are composed of 24 FDA-approved drugs and 14 experimental drugs.[14] The synergy score of each drug pair was calculated by using the Combenefit tool.[26] According to a previous study,[22] we selected 10 as a threshold to classify the drug pair−cell-line triplets. The triplets with a synergistic score higher than 10 are recognized as positive, and those less than 0 are seen as negative. After preprocessing the data, we obtained 13,243 unique triplets, consisting of 38 drugs and 31 cell lines. Moreover, the SMILES (Simplified Molecular Input Line Entry System)[27] of drugs is obtained from DrugBank.[28]

The gene expression data of cancer cell lines are obtained from Cancer Cell Line Encyclopedia (CCLE),[29] which is an independent project that makes the effort to characterize genomes, mRNA expression, and anticancer drug dose responses across cancer cell lines. The expression data is normalized through TPM (Transcripts Per Million) based on the genome-wide read counts matrix.

Framework of AttenSyn. In this section, we introduce the details of our proposed AttenSyn. The overall architecture of AttenSyn is shown in Figure 1. This network architecture mainly includes three parts: (1) Graph-based drug-embedding module, (2) attention-based pooling module, and (3) prediction module. In the graph-based drug-embedding module, first, the drug SMILES strings are transformed into molecular graphs, and simultaneously the cell-line features obtained from CCLE[29] are added to the feature matrices of the drug molecules. Then, several graph convolution network (GCN) models and LSTM models are employed to extract the multiresolution features of molecular graphs. After that, we use the attention-based pooling module to learn the interactive information between drug pairs and strengthen the representations of drug pairs. Finally, in the prediction module, we concatenated the representations of drug pairs and features of cell lines and fed them into a fully connected neural network to predict the synergy of drug pairs in certain cell lines.

Graph-Based Drug-Embedding Module. By using the open-source python package Rdkit,[30] we can convert the

https://doi.org/10.1021/acs.jcim.3c00709
J. Chem. Inf. Model. 2024, 64, 2854−2862

SMILES string into a molecular graph where the nodes are atoms and the edges are chemical bonds. Thus, we can use a graph G = (V, E) to represent a drug molecule, where V and E are the set of nodes and the set of edges, respectively. To aggregate molecular graphs with cell-line information, we simply add the cell-line vector to the nodes' features as follows:

$$h_i^0 = x_i + \mathrm{MLP}(R_{\text{cell line}}) \tag{1}$$

where x_i denotes the feature vector of node i and $R_{\text{cell line}}$ is the cell-line vector obtained by CCLE.

In order to obtain the representation of chemical substructures, we employ a GNN module that uses a chemical graph structure as the input and updates vector embeddings of each atom from its neighbors. Therefore, the updated feature vector of each atom can represent chemical substructures. The GCN operator we used can be formulated as follows:

$$H^{l+1} = \sigma(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^lW^l) \tag{2}$$

where $\tilde{A} = A + I$, I is the identity matrix, and A is the adjacency matrix. $\tilde{D}$ is the degree matrix of $\tilde{A}$, which is calculated by $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. W^l is a layer-specific trainable weight matrix. H^l represents the learned representations by the lth layer and $H^0 = [h_1^0, h_2^0, h_3^0, ..., h_i^0, ...]$. σ denotes the nonlinear activation function.

In order to get multiresolution information on molecular graphs, inspired by MR-GNN,[31] we use LSTM to extract the graph's multiresolution local features. Several LSTM models are used to aggregate the features of multiple GCN layers. Specifically, the LSTM sequentially receives the output of each GNN layer with a receptive field from small to large as the input. The LSTM can be formulated as

$$S^{l+1} = \mathrm{LSTM}(S^l, H^l) \tag{3}$$

where S^{l+1} is the $(l+1)$th hidden vector of LSTM.

Attention-Based Pooling Module. To improve the performance of the proposed model, we designed an attention-mechanism-based pooling to learn better interactive information on drug pairs and strengthen the representations of drugs. The use of attention-based pooling helps the proposed model consider which substructures in the chemical are more important for the prediction of synergistic drug combinations. As shown in Figure 2, the attention-based pooling module was used to assign each substructure of the drug a score, and weighted sum all nodes' embeddings to get graph-level representations. By using our designed attention-based pooling module, we can not only get the interactive information between drug pairs but also identify the important chemical substructures of drugs. We calculate the attention scores as follows:

$$A_x = \tanh(H_x^lW_k(H_y^lW_q)^{\mathrm{T}}) \tag{4}$$

$$A_y = \tanh(H_y^lW_k(H_x^lW_q)^{\mathrm{T}}) \tag{5}$$

$$a_x = \mathrm{softmax}\left(\sum_{j=1}^{M} A_{xi,j}\right) \tag{6}$$

$$a_y = \mathrm{softmax}\left(\sum_{j=1}^{N} A_{yi,j}\right) \tag{7}$$

where H_x^l and H_y^l are the graph-embedding matrices of drug$_x$ and drug$_y$ respectively, taken from the last GCN layer. And a_x, a_y are the attention scores for the drug pair. Then we weighted sum all the nodes' vectors according to the attention scores to get the final graph-level representations:

$$g_x = \mathrm{multiply}(a_x, H_xW_v) \tag{8}$$

$$g_y = \mathrm{multiply}(a_y, H_yW_v) \tag{9}$$

where H_x and H_y are the embedding matrices of the drug pair in the last GCN layer. For the output of LSTM, we calculate the final representation in the same way, and the formula is as follows:

$$s_x, s_y = \text{attention-based pooling}(S_x, S_y) \tag{10}$$

where S_x and S_y denote the embedding matrices of the drug pair in the last LSTM layer and the s_x and s_y are the graph-level feature vectors of the two drugs captured by LSTM.

Prediction Module. In this module, we first integrate all the features of two drugs with the feature vector of the cell line and then make predictions using a multilayer perceptron (MLP). Specifically, we concatenate the features of the drug pair from GCN and LSTM $g_x g_y s_x s_y$ and cell-line feature vector $R_{\text{cell line}}$, and then the MLP is used for classification.

$$p_i = \mathrm{softmax}\left(\mathrm{MLP}\left(g_x \| g_y \| s_x \| s_y \| \mathrm{MLP}(R_{\text{cell line}})\right)\right) \tag{11}$$

where $\|$ is the concatenation operator.

Our model is optimized by minimizing the cross entropy loss function:

$$\text{cross entropy} = \frac{1}{N}\sum_{i=1}^{N} -[y_i \cdot \log(p_i) + (1 - y_i)\cdot \log(1 - p_i)] \tag{12}$$

where N is the total number of samples in the training set, y_i is the label of sample i, and p_i denotes the probability of identifying this drug pair to the synergistic combination by our model.

Metrics. For the anticancer drug synergy prediction task, we adopted metrics including the area under the receiver operator characteristics curve (AUROC), the area under the precision−recall curve (AUPR), accuracy (ACC), balanced accuracy (BACC), precision (PREC), true positive rate (TPR), and the Cohen's Kappa value (KAPPA). These indicators are calculated by the following formula:

$$\begin{cases} \mathrm{ACC} = \dfrac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{TN} + \mathrm{FP}} \\ \mathrm{BACC} = \dfrac{\mathrm{TPR} + \mathrm{TNR}}{2} \\ \mathrm{TPR} = \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \\ \mathrm{TNR} = \dfrac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \\ \mathrm{PREC} = \dfrac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \\ \mathrm{KAPPA} = \dfrac{p_o - p_e}{1 - p_e} \end{cases} \tag{13}$$

https://doi.org/10.1021/acs.jcim.3c00709
J. Chem. Inf. Model. 2024, 64, 2854−2862

Table 1. Performances of Our Method and Existing Methods on Benchmark Datasets

Methods	AUROC	AUPR	ACC	BACC	PREC	TPR	KAPPA
AttenSyn	0.92 ± 0.01	0.91 ± 0.01	0.84 ± 0.01	0.84 ± 0.02	0.83 ± 0.02	0.82 ± 0.03	0.67 ± 0.03
DTSyn	0.89 ± 0.01	0.87 ± 0.01	0.81 ± 0.01	0.81 ± 0.02	0.84 ± 0.02	0.74 ± 0.05	0.61 ± 0.03
MR-GNN	0.90 ± 0.01	0.90 ± 0.01	0.82 ± 0.01	0.82 ± 0.01	0.81 ± 0.02	0.80 ± 0.03	0.65 ± 0.02
DeepSynergy	0.72 ± 0.01	0.77 ± 0.03	0.72 ± 0.01	0.72 ± 0.01	0.73 ± 0.05	0.64 ± 0.02	0.43 ± 0.02
RF	0.74 ± 0.03	0.73 ± 0.03	0.67 ± 0.01	0.67 ± 0.02	0.70 ± 0.07	0.59 ± 0.03	0.35 ± 0.04
Adaboost	0.74 ± 0.02	0.72 ± 0.03	0.75 ± 0.02	0.66 ± 0.02	0.63 ± 0.08	0.69 ± 0.08	0.32 ± 0.04
SVM	0.68 ± 0.05	0.65 ± 0.06	0.62 ± 0.05	0.62 ± 0.05	0.59 ± 0.05	0.66 ± 0.06	0.25 ± 0.09
MLP	0.84 ± 0.01	0.82 ± 0.01	0.76 ± 0.01	0.75 ± 0.01	0.75 ± 0.01	0.71 ± 0.01	0.50 ± 0.02
Elastic net	0.68 ± 0.08	0.67 ± 0.07	0.63 ± 0.07	0.63 ± 0.07	0.61 ± 0.08	0.62 ± 0.07	0.27 ± 0.14

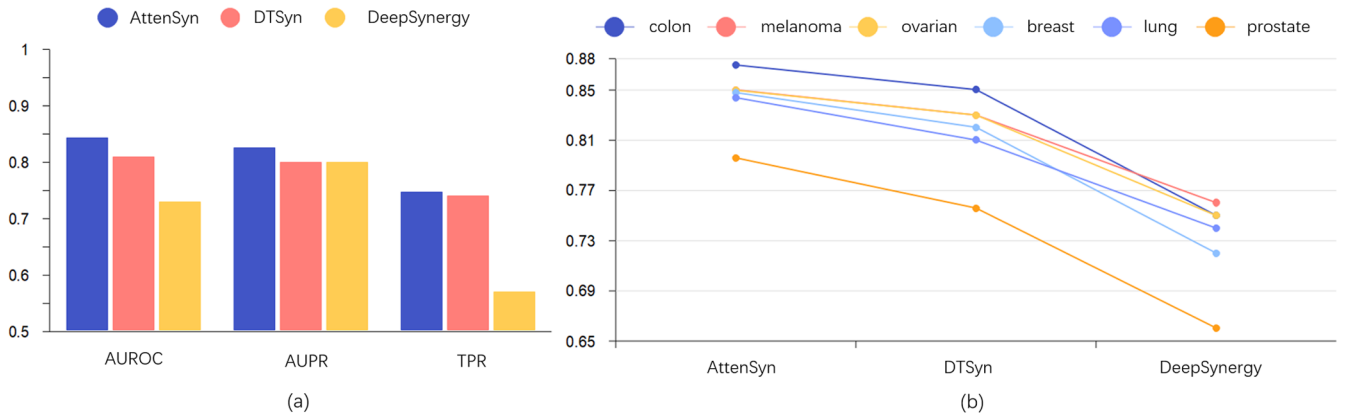


Figure 3. The performances of our method and the other two methods on the leave-tumor-out cross validation task. (a) AUROC, AUPR, and TPR of all comparing methods on the leave-tumor-out cross validation task; (b) AUROC scores of our AttenSyn and the other two methods on each tumor type.

Among these, TP (true positive), FN (false negative), TN (true negative), and FP (false positive) represent the number of identifying synergistic drug combinations as synergistic drug combinations, identifying synergistic drug combinations as antagonistic drug combinations, identifying antagonistic drug combination as antagonistic drug combination, and identifying antagonistic drug combination as synergistic drug combination by the model, respectively. p_o is the empirical probability of agreement on the label assigned to any sample, and p_e is the expected agreement when both annotators assign labels randomly.

ACC describes how the model differs across two classes and is useful in binary classification.[32] BACC and KAPPA are two indicators that take into account the predictive power of synergistic drug combinations and antagonistic drug combinations and are widely used in unbalanced datasets. PREC measures the prediction accuracy of drug pairs that are predicted as synergistic drug combinations. TPR and TNR represent the accuracy of the prediction results of the predictor for the positive and negative samples, respectively. Generally speaking, the higher the above indicators, the better the prediction ability of the model.

RESULTS AND DISCUSSION

Comparison with Existing Methods on Benchmark Datasets. To evaluate the effectiveness of the proposed AttenSyn, we compare it with several existing methods, including machine-learning-based methods (i.e., random forest (RF), Support Vector Machines (SVM), Multilayer Perceptron (MLP), Adaboost, and Elastic net) and deep-learning methods (i.e., DTSyn,[22] MR-GNN,[31] and DeepSynergy[14]), by five fold cross validation on the benchmark datasets. The detailed comparative results are illustrated in Table 1, and the best results are shown in bold. As shown in Table 1, we can see that our proposed AttenSyn achieved better performance than competing methods in terms of AUROC, AUPR, ACC, BACC, TPR, and KAPPA. Specifically, our AttenSyn achieved AUROC, AUPR, ACC, BACC, PREC, TPR, and KAPPA of 0.92, 0.91, 0.84, 0.84, 0.83, 0.82, and 0.67, respectively. We note that the other two molecular-graph-based deep-learning methods, DTSyn and MR-GNN, also achieved remarkable performances, which follow our method closely and outperform much more than other methods.

To further demonstrate the good performance of our AttenSyn, we use the leave-tumor-out cross validation strategy to evaluate our method and the other two state-of-the-art deep-learning methods (i.e., DTSyn and DeepSynergy). More precisely, to make sure that the model cannot see any gene expression information on a specific type of tumor in the training process, we exclude all the cancer cell lines belonging to the specific tumor from the training set. We repeat the process and iteratively use the excluded cancer cell lines as the validation set and the remaining samples as the training set to train and evaluate the model. Figure 3a shows the comparison result of AttenSyn and the other two deep-learning-based methods on the leave-tumor-out cross validation task. It can be seen from Figure 3a that our proposed AttenSyn achieved the best on AUROC, AUPR, and TPR. Figure 3b shows the AUROC score of AttenSyn and the other two methods on each tumor type. As shown in Figure 3b, we can see that AttenSyn has the best AUROC score over all the six tumor types, indicating that our AttenSyn has the potential to predict synergistic drug combinations across various tumor types. We

https://doi.org/10.1021/acs.jcim.3c00709
J. Chem. Inf. Model. 2024, 64, 2854−2862

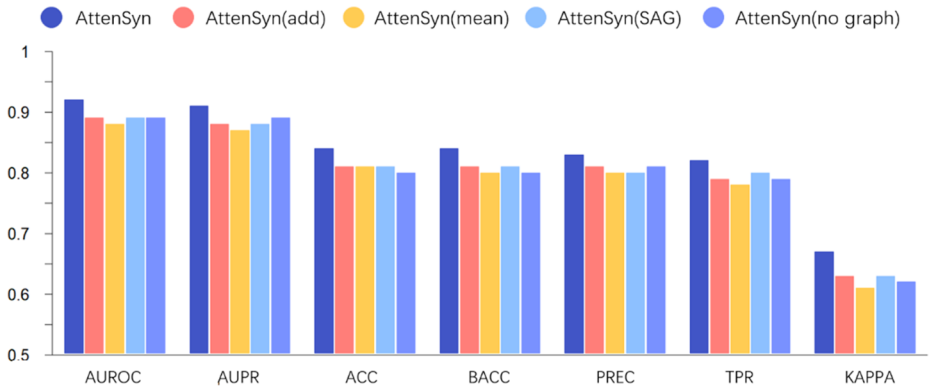


Figure 4. Performances of our proposed AttenSyn and its variants.

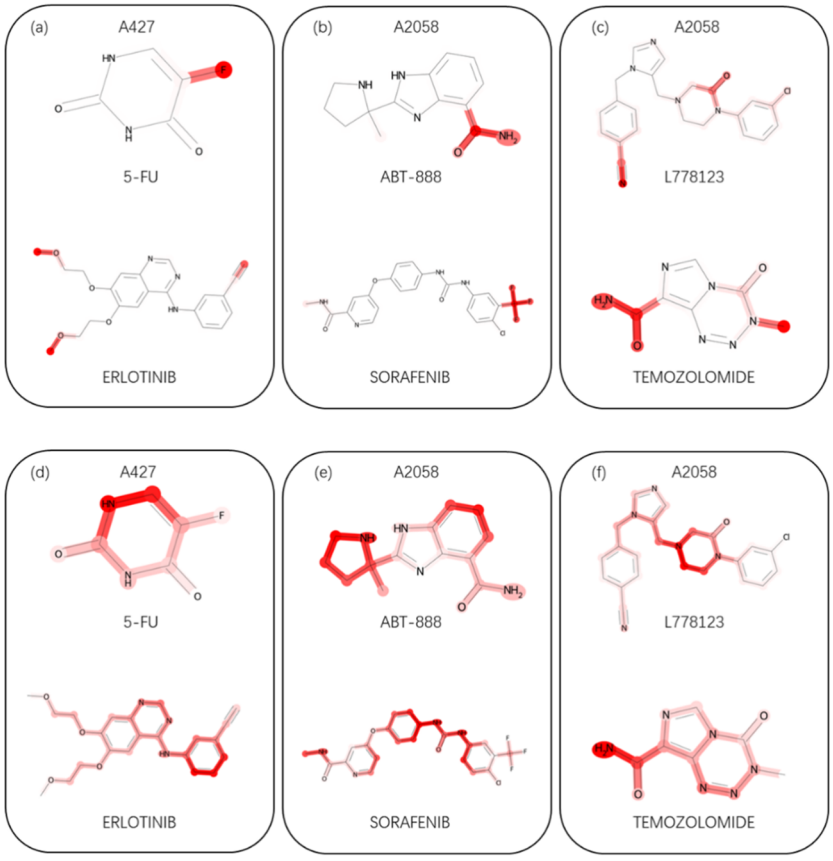


Figure 5. Visualization results of three randomly selected drug pairs. Panels (a)−(c) show the visualization of attention scores from our model in the three-drug pairs after the training process; panels (d)−(f) show the visualization of attention scores of the three-drug pairs before the training process. The darker color represents a more important substructure.

also conduct the experiment under cold start setting and our AttenSyn achieved the best performance compared with other methods. The results can be seen in Supplementary Figure 1.

Ablation Study. To investigate the effect of the attention-based pooling module and graph-based drug-embedding module on our model performance, we consider the following variants of AttenSyn: (1) the proposed AttenSyn; (2) AttenSyn (add); (3) AttenSyn (mean); (4) AttenSyn (SAG); (5) AttenSyn(no graph). Specifically, the AttenSyn (add) uses a global add pooling method instead of attention-based pooling in our original AttenSyn. The AttenSyn (mean) uses a global mean pooling method instead of attention-based pooling. And the AttenSyn (SAG) calculates self-attention scores in the way introduced by SAGPooling[33] to update its nodes' embeddings and then add them to get the graph-level representations. The AttenSyn(no graph) removed the graph-based drug-embedding module.

As can be seen in Figure 4, the performance of our original AttenSyn is better compared with other variants of AttenSyn. Moreover, from the results in Figure 4, we can see that AttenSyn (add) and AttenSyn (SAG) are slightly superior to the results of AttenSyn (mean). This may be because, for molecular graphs, the global mean pooling method treats each substructure as equally important and simply averages all nodes' embeddings. Specifically, SAGPooling uses a self-attention mechanism to calculate the score of each

https://doi.org/10.1021/acs.jcim.3c00709
J. Chem. Inf. Model. 2024, 64, 2854−2862

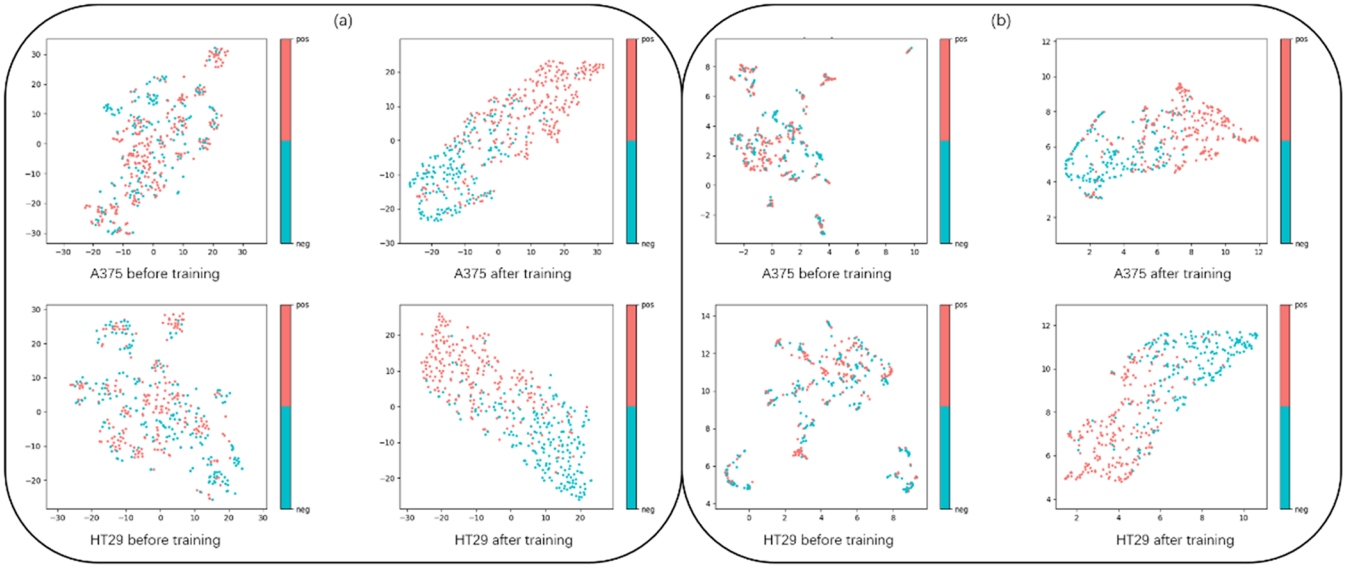


Figure 6. Visualization of the feature space distribution of our model. Panel (a) represents t-SNE visualization results of our model with and without the training process on A375 and HT29; panel (b) represents UMAP visualization results of our model with and without the training process on A375 and HT29.

substructure and weighted sum all the nodes' embeddings according to the attention scores to get the final graph-level representation. However, there is no significant difference in all seven metrics between AttenSyn (add) and AttenSyn (SAG). The reason might be that the SAGPooling method uses only a single drug's molecular graph to get the attention scores of each substructure instead of using interactive information on drug pairs to calculate the importance of each substructure. Meanwhile, our proposed attention-based pooling module uses interactive information on drug pairs and thus can achieve better performance than the other pooling strategies. In addition, our proposed AttenSyn is superior to the AttenSyn-(no graph), proving the effectiveness of graph-based drug-embedding module.

Visualization of Important Substructures Detected by Our Model. Deep-learning-based models are often regarded as "black box", and the lack of model interpretability limits their further application in real scenarios of many areas, especially in computer-aided drug discovery. To overcome the black box problem and explore which substructures within the drug pairs provide the most significant contribution to synergistic drug combinations prediction, we visualized the most essential substructures for drug pairs through the attention mechanism of our model. Specifically, we use the attention score calculated by formulas 6 and 7 to represent the importance of corresponding substructures and visualize the scores in different colors. Figure 5a−c shows the visualization results of three randomly selected example drug pairs, where the darker color denotes the more important substructure. One of the chemical structures detected by our model is the amide group, which plays a key role in the composition of biomolecules, including many clinically approved drugs.[34] Amides are prevalent in medicinally important compounds not only because they are particularly stable but also because they are polar, which allows amide-containing drugs to interact with biological receptors and enzymes.[35] This result demonstrates that our model can provide good interpretability.

To further explore the changes in attention scores of substructures during the training process, we also visualized the attention score's distribution of drug pairs before model training. As shown in Figure 5d−f, the attention scores before model training are distributed more uniformly, which indicated that the model can not pay attention to the important structures. However, as the training goes on, some specific structures have been considered more important than others by the model.

Feature Representation and Visualization by Dimension Reduction. To further interpret how deep learning works during the training process from the feature analysis and intuitively show the feature learning ability of the proposed AttenSyn, we visualized the embeddings of drug combinations in two cell lines (i.e., A375 and HT29). Specifically, we reduce the embedding space of drug pairs extracted from our model with and without the training process to a two-dimensional space by using t-SNE[36] and UMAP,[37] respectively, as in Figure 6. In each subfigure of Figure 6, each point represents a drug pair, and different colors are used to distinguish the synergistic drug combination and antagonistic drug combination classes. The more distinguishable the points under different categories are, the better the classification effect is. As shown in Figure 6a, by the dimension reduction of t-SNE, the samples of two classes are distributed more clearly in the feature space of the trained model compared to the model without the training process, indicating that our model can capture discriminative and high-quality features from different classes samples. There exist similar results for the model by the dimension reduction of UMAP. From Figure 6b, the model with the training process learns and achieves more distinguishable features compared to the model without the training process.

CONCLUSION

In this study, we developed a novel attention-based deep graph neural network named AttenSyn to predict the synergy of anticancer drug combinations, which is a crucial step for rapid virtual drug screening and drug development. Specifically, we first generate molecular graphs of drugs and employ the graph-based drug-embedding module to extract structural information on drug pairs, respectively. After that, the attention-based

https://doi.org/10.1021/acs.jcim.3c00709
J. Chem. Inf. Model. 2024, 64, 2854−2862

pooling module is designed to learn better interactive information and strengthen the representations of drug pairs. Comprehensive experiments conducted on the benchmark datasets show that the proposed method achieves a better predictive performance than the comparative methods. Moreover, to overcome the limitations of the "black box" in deep-learning-based models, we explored what our model learns during the training process in both discovering the crucial substructures in drugs and conducting feature analysis, which provides good interpretability of our model and biological insights for understanding the drug synergy mechanism.

However, there are still some drawbacks to our model. For example, biological networks have already proven their effectiveness in drug synergy prediction.[15,38−41] We use only the molecular structure information and cell-line features, but not extra information such as biological network information for prediction. In the future, we will consider introducing the biological network to improve the performance of anticancer synergistic drug combinations prediction.

■ ASSOCIATED CONTENT

Data Availability Statement

The data and code for this study can be found in a GitHub repository accompanying this manuscript: https://github.com/badguyisme/SynPred.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c00709.

Experimental settings, performance under cold start setting, and the distributions of predictive probability (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Leyi Wei − *School of Software, Shandong University, Jinan 250101, China; Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan 250101, China;* orcid.org/0000-0003-1444-190X; Email: weileyi@sdu.edu.cn

Authors

Tianshuo Wang − *School of Software, Shandong University, Jinan 250101, China; Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan 250101, China*

Ruheng Wang − *School of Software, Shandong University, Jinan 250101, China; Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), Shandong University, Jinan 250101, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.3c00709

Author Contributions

T.W. conceived the basic idea and designed the framework. T.W. and R.W. performed the experiments. R.W. and T.W. wrote the manuscript. L.W. revised the manuscript. R.W. conducted the visualization of the experimental results.

Funding

The work was supported by the Natural Science Foundation of China (Nos. 62322112 and 62071278).

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Giles, T. D; Weber, M. A; Basile, J.; Gradman, A. H; Bharucha, D. B; Chen, W.; Pattathil, M. Efficacy and safety of nebivolol and valsartan as fixed-dose combination in hypertension: a randomised, multicentre study. *Lancet* **2014**, *383* (9932), 1889−1898.

(2) Zheng, W.; Sun, W.; Simeonov, A. Drug repurposing screens and synergistic drug-combinations for infectious diseases. *British journal of pharmacology* **2018**, *175* (2), 181−191.

(3) Kim, Y.; Zheng, S.; Tang, J.; Jim Zheng, W.; Li, Z.; Jiang, X. Anticancer drug synergy prediction in understudied tissues using transfer learning. *Journal of the American Medical Informatics Association* **2021**, *28* (1), 42−51.

(4) Chen, X.; Yan, C. C.; Zhang, X.; Zhang, X.; Dai, F.; Yin, J.; Zhang, Y. Drug-target interaction prediction: databases, web servers and computational models. *Briefings in bioinformatics* **2016**, *17* (4), 696−712.

(5) Chen, X.; Guan, N.-N.; Sun, Y.-Z.; Li, J.-Q.; Qu, J. MicroRNA-small molecule association identification: from experimental results to computational models. *Briefings in Bioinformatics* **2020**, *21* (1), 47−61.

(6) Wang, C.-C.; Zhao, Y.; Chen, X. Drug-pathway association prediction: from experimental results to computational models. *Briefings in Bioinformatics* **2021**, *22* (3), bbaa061.

(7) Sałat, R.; Sałat, K. The application of support vector regression for prediction of the antiallodynic effect of drug combinations in the mouse model of streptozocin-induced diabetic neuropathy. *Computer methods and programs in biomedicine* **2013**, *111* (2), 330−337.

(8) Qi, Y. Random forest for bioinformatics. In *Ensemble machine learning*; Springer: 2012; pp 307−323.

(9) Liu, H.; Zhang, W.; Nie, L.; Ding, X.; Luo, J.; Zou, L. Predicting effective drug combinations using gradient tree boosting based on features extracted from drug-protein heterogeneous network. *BMC Bioinformatics* **2019**, *20* (1), 645.

(10) Pivetta, T.; Isaia, F.; Trudu, F.; Pani, A.; Manca, M.; Perra, D.; Amato, F.; Havel, J. Development and validation of a general approach to predict and quantify the synergism of anti-cancer drugs using experimental design and artificial neural networks. *Talanta* **2013**, *115*, 84−93.

(11) Zhang, C.; Yan, G. Synergistic drug combinations prediction by integrating pharmacological data. *Synthetic and systems biotechnology* **2019**, *4* (1), 67−72.

(12) Janizek, J. D.; Celik, S.; Lee, S.-I. Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. *BioRxiv*, 2018, 331769 (submitted 6/29/2023) (accessed 7/16/2023).

(13) Chen, X.; Ren, B.; Chen, M.; Wang, Q.; Zhang, L.; Yan, G. NLLSS: predicting synergistic drug combinations based on semi-supervised learning. *PLoS computational biology* **2016**, *12* (7), No. e1004975.

(14) Preuer, K.; Lewis, R. P.; Hochreiter, S.; Bender, A.; Bulusu, K. C.; Klambauer, G. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* **2018**, *34* (9), 1538−1546.

(15) Liu, Q.; Xie, L. TranSynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS computational biology* **2021**, *17* (2), No. e1008653.

(16) Kuru, H. I.; Tastan, O.; Cicek, A. E. MatchMaker: a deep learning framework for drug synergy prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2022**, *19* (4), 2334−2344.

(17) Su, R.; Huang, Y.; Zhang, D.-g.; Xiao, G.; Wei, L. SRDFM: Siamese Response Deep Factorization Machine to improve anti-cancer drug recommendation. *Briefings in Bioinformatics* **2022**, *23* (2), bbab534.

(18) Lin, W.; Wu, L.; Zhang, Y.; Wen, Y.; Yan, B.; Dai, C.; Liu, K.; He, S.; Bo, X. An enhanced cascade-based deep forest model for drug combination prediction. *Briefings in Bioinformatics* **2022**, *23* (2), bbab562.

(19) Li, T.-H.; Wang, C.-C.; Zhang, L.; Chen, X. SNRMPACDC: computational model focused on Siamese network and random matrix projection for anticancer synergistic drug combination prediction. *Briefings in Bioinformatics* **2023**, *24* (1), bbac503.
(20) Wang, J.; Liu, X.; Shen, S.; Deng, L.; Liu, H. DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics* **2022**, *23*, (1), DOI: 10.1093/bib/bbab390.
(21) Jin, W.; Stokes, J. M.; Eastman, R. T.; Itkin, Z.; Zakharov, A. V.; Collins, J. J.; Jaakkola, T. S.; Barzilay, R. Deep learning identifies synergistic drug combinations for treating COVID-19. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (39), No. e2105070118.
(22) Hu, J.; Gao, J.; Fang, X.; Liu, Z.; Wang, F.; Huang, W.; Wu, H.; Zhao, G. DTSyn: a dual-transformer-based neural network to predict synergistic drug combinations. *bioRxiv* 2022 (submitted 6/29/2023) (accessed 7/16/2023).
(23) Nyamabo, A. K.; Yu, H.; Shi, J.-Y. SSI-DDI: substructure-substructure interactions for drug-drug interaction prediction. *Briefings in Bioinformatics* **2021**, *22* (6), bbab133.
(24) Deac, A.; Huang, Y.-H.; Veličković, P.; Liò, P.; Tang, J. Drug-drug adverse effect prediction with graph co-attention. *arXiv*, 1905.00534, 2019 (submitted 6/29/2023) (accessed 7/16/2023).
(25) O'Neil, J.; Benita, Y.; Feldman, I.; Chenard, M.; Roberts, B.; Liu, Y.; Li, J.; Kral, A.; Lejnine, S.; Loboda, A.; Arthur, W.; Cristescu, R.; Haines, B. B.; Winter, C.; Zhang, T.; Bloecher, A.; Shumway, S. D. An unbiased oncology compound screen to identify novel combination strategies. *Molecular cancer therapeutics* **2016**, *15* (6), 1155−1162.
(26) Di Veroli, G. Y.; Fornari, C.; Wang, D.; Mollard, S.; Bramhall, J. L.; Richards, F. M.; Jodrell, D. I. Combenefit: an interactive platform for the analysis and visualization of drug combinations. *Bioinformatics* **2016**, *32* (18), 2866−2868.
(27) Weininger, D. J. J. o. c. i. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* **1988**, *28* (1), 31−36.
(28) Wishart, D. S; Feunang, Y. D; Guo, A. C; Lo, E. J; Marcu, A.; Grant, J. R; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* **2018**, *46* (D1), D1074−D1082.
(29) Ghandi, M.; Huang, F. W.; Jane-Valbuena, J.; Kryukov, G. V.; Lo, C. C.; McDonald, E. R.; Barretina, J.; Gelfand, E. T.; Bielski, C. M.; Li, H.; Hu, K.; Andreev-Drakhlin, A. Y.; Kim, J.; Hess, J. M.; Haas, B. J.; Aguet, F.; Weir, B. A.; Rothberg, M. V.; Paolella, B. R.; Lawrence, M. S.; Akbani, R.; Lu, Y.; Tiv, H. L.; Gokhale, P. C.; de Weck, A.; Mansour, A. A.; Oh, C.; Shih, J.; Hadi, K.; Rosen, Y.; Bistline, J.; Venkatesan, K.; Reddy, A.; Sonkin, D.; Liu, M.; Lehar, J.; Korn, J. M.; Porter, D. A.; Jones, M. D.; Golji, J.; Caponigro, G.; Taylor, J. E.; Dunning, C. M.; Creech, A. L.; Warren, A. C.; McFarland, J. M.; Zamanighomi, M.; Kauffmann, A.; Stransky, N.; Imielinski, M.; Maruvka, Y. E.; Cherniack, A. D.; Tsherniak, A.; Vazquez, F.; Jaffe, J. D.; Lane, A. A.; Weinstock, D. M.; Johannessen, C. M.; Morrissey, M. P.; Stegmeier, F.; Schlegel, R.; Hahn, W. C.; Getz, G.; Mills, G. B.; Boehm, J. S.; Golub, T. R.; Garraway, L. A.; Sellers, W. R. Next-generation characterization of the cancer cell line encyclopedia. *Nature* **2019**, *569* (7757), 503−508.
(30) Landrum, G. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. https://www.rdkit.org/RDKit_Overview.pdf, 2013.
(31) Xu, N.; Wang, P.; Chen, L.; Tao, J.; Zhao, J. Mr-gnn: Multi-resolution and dual graph neural network for predicting structured entity interactions. *arXiv*, 1905.09558, 2019 (submitted 6/29/2023) (accessed 7/16/2023).
(32) Wang, R.; Jin, J.; Zou, Q.; Nakai, K.; Wei, L. Predicting protein-peptide binding residues via interpretable deep learning. *Bioinformatics* **2022**, *38* (13), 3351−3360.
(33) Lee, J.; Lee, I.; Kang, J. Self-attention graph pooling; In *International Conference on Machine Learning, PMLR: 2019*; pp 3734−3743.
(34) Kumari, S.; Carmona, A. V.; Tiwari, A. K.; Trippier, P. C. Amide bond bioisosteres: Strategies, synthesis, and successes. *Journal of medicinal chemistry* **2020**, *63* (21), 12290−12358.
(35) Clayden, J. *Fluorinated compounds present opportunities for drug discovery*. Nature Publishing Group: London, 2019.
(36) Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **2008**, *9*(11).
(37) McInnes, L.; Healy, J.; Melville, J., Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv*, 1802.03426, 2018 (submitted 6/29/2023) (accessed 7/16/2023).
(38) Yang, J.; Xu, Z.; Wu, W. K. K.; Chu, Q.; Zhang, Q. GraphSynergy: a network-inspired deep learning model for anticancer drug combination prediction. *Journal of the American Medical Informatics Association* **2021**, *28* (11), 2336−2345.
(39) Meng, F.; Li, F.; Liu, J.-X.; Shang, J.; Liu, X.; Li, Y. NEXGB: A Network Embedding Framework for Anticancer Drug Combination Prediction. *International Journal of Molecular Sciences* **2022**, *23* (17), 9838.
(40) Jiang, P.; Huang, S.; Fu, Z.; Sun, Z.; Lakowski, T. M.; Hu, P. Deep graph embedding for prioritizing synergistic anticancer drug combinations. *Computational and structural biotechnology journal* **2020**, *18*, 427−438.
(41) Zhang, P.; Tu, S. A knowledge graph embedding-based method for predicting the synergistic effects of drug combinations. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; IEEE: 2022; pp 1974−1981.