Predicting Drug Combination Synergies Based On Transformer Networks

Yiding Zhang

***Abstract*. -- As research into tumours continues, a large number of targeted drugs for the treatment of tumours are being devised. However, most of the drugs currently available for cancer treatment only target a single type of cancer, and single-use drugs have low drug utilisation and are prone to drug resistance. Combinations can be an excellent way to address these issues with single medicines, but because there are so many different kinds of anti-cancer medications on the market, it is inevitable that using just experimental approaches to screen combinations will be ineffective. Deep learning techniques are required to forecast the effects of combination medications, which are subsequently tested experimentally, in order to solve this issue. In this paper, we propose to use a model of transformer arithmetic to predict the effect of drug combinations on cell lines. After being fed into a multi-layer feed-forward neural network, information on the drug's molecular structure and the proteomics of the cancer cell line is used as input to the model, and this feature information is then used as input to the TRANSFORMER to predict the action fraction of the drug on a particular cancer cell line. And our model is compared with other machine learning and deep learning models. On the independent test set, our model outperforms the other models.**

***Index Terms*—Drug combination synergies, transformer, cancer, attention mechanism, deep learning**

I. INTRODUCTION

The right combination of drugs can improve the efficacy of drugs, reduce their toxic effects and reduce drug resistance[1]. In some practical cases, the combination of older drugs has been effective in the treatment of multiple diseases. For example, in the treatment of triple-negative breast cancer, lapatinib or rapamycin alone have little effect, but their combination significantly increases the rate of apoptosis in triple-negative breast cancer cells to some extent[2].

It is clearly impossible to screen the effects of combinations of drugs one by one by experimental means, and with a large number of drugs and a large number of cancer cell lines, these would correspond to tens of millions of combinations, and it would be impossible to complete all of them. Therefore, corresponding computational models need to be proposed to solve these problems[3-4].

A large number of models have been proposed to accomplish the corresponding task[5]. Among them, Iwata et al used a logistic regression algorithm to predict effective drug combinations from target protein information as well as ACT drug coding, and to reduce overfitting by L1 regularisation[6]. The PEA model predicts the effect of a combination of drugs by probability. Given a pair of drugs, PEA calculates the similarity characteristics of the drugs and uses a Bayesian network to calculate the likelihood ratio to obtain the similarity in probability of a known combination of drugs[7]. The SyDRa model predicts drug combinations by means of three random forest models[8]. These early machine learning models, though typically simpler, less computationally intensive, and with some interpretability, correspond to models that are less predictively accurate than deep learning models, are more reliant on input data, and improving model performance is a more challenging task. among the existing deep learning techniques for predicting drug synergy. DeepSynergy predicts drug combination scores by simply combining drug and cell line features in a multi-layer fully connected network[9]. DeepDDS uses the smiles information of the drug to learn the drug features, but the final training to get the drug features and cell line features is still through a simple feed-forward neural network to get the drug action score[10]. Transynergy also uses a transformer to predict drug combination effects, while adding separate proteomic information as input[11]. but Transynergy uses multiple transformer modules, resulting in a more computationally intensive model and the risk of overfitting, leading to poor model performance on independent test sets.

In this paper, we propose a model to predict drug combination synergies based on the TRANSFORMER. In contrast to other deep learning models, our model takes the drug and cancer cell line features and learns the correlation between the drug combination and the cell line through the attention mechanism in the TRANSFORMER before combining these features to predict the drug action score. Our model outperforms DTSyn on different test sets and is simpler and faster to compute than DTSyn, which makes use of multiple attention modules.

II. MATERIALS AND METHODS

A. Data

We use the smiles of the drugs as input to the model. The smiles of the drugs are obtained from DrugBank. The drug's smiles are converted into an isomorphic map using RDKit[12], where the atoms act as nodes and the chemical bonds as edges. The whole isomorphic map is then fed into the model.

Gene expression in cancer cell lines from TCGA. The TCGA includes multi-omics data from 33 different patients with different cancer types, mainly including gene expression, mutation, DNA methylation, and copy number change data.

Drug combination sensitivity data is obtained from the Large Scale Cancer Screening Dataset, a biochemical approach to assess whether a drug combination has an effect

Manuscript received November 26, 2022.
Yiding Zhang, Tiangong University/School of Computer Science and Technology, Tianjin, China

on a cell line. The Complexation Tool's 4 to 4 dose-response matrix was used to calculate a specific drug combination action score, with an action score greater than 4 denoting a positive synergy between the two drugs and less than 4 denoting no positive synergy. More than 20,000 clock drug combinations' final action scores on 39 cell lines were obtained and divided into three sets: a training set, a test set, and a validation set in the ratio 8:1:1.

B. Models

Fig.1 illustrates the main framework of our model, taking an end-to-end learning approach. Molecular maps of drugs and gene expression of cell lines are used as input. As the molecular map of the drug is not directly input as features into the subsequent model, the features of the drug are extracted using a graph attention network(GAT). Cell lineage characteristics are mapped by MLP to the same feature dimension as the drug feature dimension. The drug features and cell line features after processing are encoded as the same class of features through transforms to learn the mechanism of attention between the drug and the cell line. Finally, drug characteristics were combined with cell line characteristics to predict drug synergy.

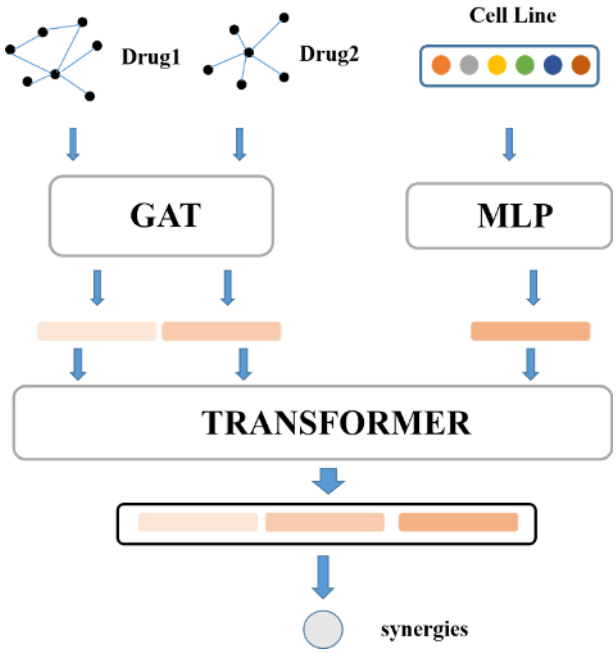

Fig.1. Pipeline of Model.

C. Drug characterisation based on GAT

We used RDKit to construct the graph structure of the drug, with the nodes of the graph being atoms and the edges of the graph being chemical bonds. This results in a graph *G(V,E)* corresponding to each drug, where *V* is a combination of nodes, here representing the set of atoms and atomic features, and E represents the connectivity of nodes in *G(V,E)*, represented by the binary group *(u, v)*. The initial features of the nodes in each graph are obtained through DeepChem.

Drug features are learned using graph attention networks, which aggregate node and neighbor node information using a multi-headed self-attentive mechanism. After learning, the node features will contain local structural information, and eventually the set of node features V will contain the drug's molecular structure information. The calculation process is as follows:

$$h_i' = ||_{k=1}^{K}(a_{il}^{k} W h_i + \sum_{j \in N(j)} a_{ij}^{k} W h_j).$$

The $\|$ denotes the multi-headed attention mechanism, K is the number of attetion heads, W is the optimizable parameter matrix, N represents the number of first-order neighbour nodes, and h is the node feature vector. The attention factor a_{ij} is calculated as follows:

$$a_{ij} = \frac{exp(ReLu(C^T[Wh_i||Wh_j]))}{\sum_{m \in N(i)} exp(ReLu(C^T[Wh_i||Wh_m]))}.$$

where *ReLu* is the non-linear activation function and C is the learnable weight vector.

D. MLP-based cell line

The cell line representations were dimensionally mapped by an MLP to correspond to the drug representation's dimensionality, and a residual network was added to avoid over-fitting. Each cancer cell line has its own genetic signature. We extracted single-cell sequencing data from 39 cancer cells in the TCGA database and used it as a characterisation of cancer cell lines after noise reduction.

E. Drug and cell line combination characteristics based on transformers

A deep learning model called the transformer makes use of attentional processes[14]. In order to acquire a feature vector after a multi-layer self-attentive network, we primarily use the encoder section of the transformer to input two drug and cell lineage vectors as a sequence into the encoder in the transformer. The attentional mechanism is calculated as follows:

First calculate the similarity coefficient between Q and K, denoted by *f*:

$$f(Q, K_i), i = 1,2, \dots, m,$$

where m denotes the number of vectors in K, Normalize the similarity in Q and K:

$$\alpha_i = \frac{e^{f(Q,K_i)}}{\sum_{j=1}^{m} f(Q,K_i)}, i = 1,2, \dots, m,$$

The vector Z after attention is obtained by weighting and summing all the values in V with respect to the computed α_i.

$$Z = \sum_{i=1}^{m} \alpha_i V_i.$$

The transformer has six layers altogether, and the attention network and residual connections work together to prevent overfitting.

III. RESULT

The hyperparameters in the model, which for our model are primarily from the MLP, GAT, and Transformer, affect the model's actual outcome as well. Different hyperparameters can greatly influence the effect of the model. The main hyperparameter settings in the model are shown in Table 1.

According to the actual experimental tests, the model works best when the number of layers in the GAT is two, with the hidden layers having dimensions of 512 and 256 respectively, and the final GAT output having a dimension of 128. There are eight headings in the GAT.The output layer has the same dimensions as the GAT output layer, and the cell lines have three hidden layers of 1024, 512, and 256 dimensions, respectively. The Transformer's orphaned layer is 128 square inches in size, and there are 8 heads.

Table I. Hyperparameter of model

Hyperparameter	Values
GAT Hidden units	[512,256]
MLP Hidden units	[1024,512,256]
Attention Head	8

We contrast the models' performance with some of the most recent models, namely the more sophisticated deep learning models and certain machine learning models, such as DeepDDS, DTsyn, DeepSynergy, etc., in order to assess how well the models performed. Additionally, we incorporate machine learning models like SVM and Random Forest.

We chose the model that performed best on the validation set to validate its performance on the test set using five-fold cross-validation, which increased the accuracy of the results the model produced.

The comparison between our model and other models for the same data set is shown in the table below.

Table II. Performance of our model compared to other methods at 5-fold cross-validation

Performance Metric	ROC AUC	PR AUC	ACC
Our model	0.85±0.01	0.85±0.01	0.83±0.01
DeepDDS	0.82±0.01	0.81±0.02	0.80±0.02
DTsyn	0.83±0.02	0.82±0.01	0.80±0.01
DeepSynergy	0.81±0.01	0.80±0.02	0.78±0.02
XGBoost	0.70±0.01	0.72±0.03	0.75±0.01
SVM	0.74±0.02	0.72±0.01	0.77±0.02
Random Forest	0.76±0.01	0.77±0.01	0.72±0.03

From the above table, it can be inferred that our model performs better on five-fold cross-validation than both the most recent deep learning models based on the graph attention mechanism and the conventional machine learning models.

The aforementioned study demonstrates that our suggested model, which requires less computing than existing sophisticated models, can accurately predict the impact of medication combinations on cancer cell lines.

IV. Conclusion

In this article, we present a fresh technique for estimating how particular cell lines will respond to various drug combinations. The molecular structure of the drug is used as an isomorphic map in the model to obtain the drug's characteristics, and the drug's characteristics obtained through GAT training are capable of accurately representing the various te of various drugs. Finally, we used a transformer module to calculate the attentional mechanism of the drug and the cell line, and further calculate the killing effect of the combination of the two drugs on the cell line.

In conclusion, our model is superior to other models in terms of prediction accuracy on five-fold cross-validation. It is important to note that although though the comparisons were conducted with other models, they were done on a single dataset, and the model has to be further validated on additional, independent datasets.

References

[1] T. -C. Chou, Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. Pharmacol Rev. 2006 Sep;58(3):621-81.

[2] T. Liu, R. Yacoub, L. D. Taliaferro-Smith, S.-Y. Sun, T. R. Graham, R. Dolan, C. Lobo, M. Tighiouart, L. Yang, A. Adams, et al, Combinatorial effects of lapatinib and rapamycin in triple-negative breast cancer cells. Molecular cancer therapeutics, 10(8):1460–1469, 2011.

[3] C. P. Goswami, L. Cheng, P. Alexander, A. Singal, and L. Li, A new drug combinatory effect prediction algorithm on the cancer cell based on gene expression and dose–response curve. CPT: pharmacometrics &systems pharmacology, 4(2):80–90, 2015..

[4] M. K. Morris, D. C. Clarke, L. C. Osimiri, and D. A. Lauffenburger, Systematic analysis of quantitative logic model ensembles predicts drug combination effects on cell signaling networks. CPT: pharmacometrics & systems pharmacology, 5(10):544–553, 2016.

[5] B. Güvenpaltun, S. Kaski, Mamitsuka H, Machine learning approaches for drug combination therapies[J]. Briefings in Bioinformatics, 2021(1).

[6] Iwata H, Sawada R, Mizutani S, et al, Large-scale prediction of beneficial drug combinations using drug efficacy and target profiles. J Chem Inf Model 2015;55(12):2705–16.

[7] Li P, Huang C, Fu Y, et al, Large-scale exploration and analysis of drug combinations. Bioinformatics 2015;31(12): 2007–16.

[8] Li X, Xu Y, Cui H, et al, Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. Artif Intell Med 2017;83: 35 – 43.

[9] K. Preuer, R. P. I. Lewis, S. Hochreiter, et al, DeepSynergy: Predicting anti-cancer drug synergy with Deep Learning[J]. Bioinformatics, 2017, 34(9).

[10] J. Wang, X. Liu, S. Shen, et al, DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations[J]. 2021.

[11] Q. Liu, L. Xie. TranSynergy: Mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations[J]. PLOS Computational Biology, 2021, 17.

[12] G. Landrum et al, Rdkit: Open-source cheminformatics. 2006.

[13] Velikovi P, Cucurull G, Casanova A, et al, Graph Attention Networks:, 10.48550/arXiv.1710.10903[P]. 2017.

[14] A. Vaswani, N. Shazeer, N. Parmar, et al, Attention Is All You Need[J]. arXiv, 2017.