

---

# CongFu: Conditional Graph Fusion for Drug Synergy Prediction

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 Drug synergy, characterized by the amplified combined effect of multiple drugs,  
2 is critically important for optimizing therapeutic outcomes. Limited data on drug  
3 synergy, arising from the vast number of possible drug combinations and testing  
4 costs, motivate the need for predictive methods. In this work, we introduce CongFu,  
5 a novel Conditional Graph Fusion Layer, designed to predict drug synergy. CongFu  
6 employs an attention mechanism and a bottleneck to extract local graph  
7 contexts and conditionally fuse graph data within a global context. Its modular  
8 architecture enables flexible replacement of layer modules, including readouts  
9 and graph encoders, facilitating customization for diverse applications. To evaluate  
10 the performance of CongFu, we conduct comprehensive experiments on four  
11 datasets, encompassing three distinct setups for drug synergy prediction. CongFu  
12 achieves state-of-the-art results on 11 out of 12 benchmark datasets, demonstrating  
13 its ability to capture intricate patterns of drug synergy. Through ablation studies,  
14 we validate the significance of individual layer components, affirming their contributions  
15 to overall predictive performance. Finally, we propose an explainability  
16 strategy for elucidating the effect of drugs on genes. By addressing the challenge of  
17 predicting drug synergy in untested drug pairs and utilizing our proposed explainability  
18 approach, CongFu opens new avenues for optimizing drug combinations  
19 and advancing personalized medicine.

## 20 1 Introduction

21 Drug combination therapy is a widely adopted approach due to its numerous advantages. Unlike  
22 monotherapy, the effect of the treatment can be significantly amplified by using a combination of  
23 drugs [1]. Furthermore, drug combinations have the potential to reduce adverse effects [2], decrease  
24 toxicity [3], and overcome drug resistance [4]. Multi-drug therapy can address complex diseases such  
25 as cancer [5, 6] or human immunodeficiency virus [7]. However, certain drug combinations may lead  
26 to unfavorable or harmful outcomes [8, 9], making it crucial to accurately predict synergistic drug  
27 pairs and potential side effects resulting from different drug interactions.

28 Historically, the discovery of drug combinations has relied on clinical trials and trial-and-error  
29 methods. These approaches are not only costly and time-consuming but can also pose risks to  
30 patients [10, 11]. Moreover, the scalability limitations of wet-lab tests restrict the screening of drug  
31 combinations [12]. However, advancements in experimental techniques have led to the development of  
32 high-throughput drug screening (HTS) [13, 14, 15], a fast and precise method that allows researchers  
33 to explore large drug combination spaces. This has resulted in a rapid increase in drug combination  
34 synergy data. Public databases like ASDCD [16] provide drug combination data and large HTS  
35 synergy studies covering numerous drugs and cancer cell lines [3]. These databases provide high-  
36 quality training data for the development of computational approaches and aid in evaluating these

37 methods for predicting novel drug combinations. However, the discrepancy between in vivo and in  
38 vitro experiments limits the effectiveness of HTS.

39 In recent years, the availability of large HTS datasets [17] has spurred the development of machine  
40 learning models for drug synergy predictions [18]. Early deep learning methods, such as DeepSynergy  
41 [19] and MatchMaker [20], utilize fully connected networks based on cell lines and drug features  
42 derived from Morgan fingerprints [21]. Subsequent models like AuDNNsynergy [22] incorporate  
43 autoencoders that leverage "copy number variation" data, gene expressions, and mutations. Other  
44 models like TranSynergy [23] adopt a transformer architecture to process a cell line and two drug  
45 feature vectors as input. DTF [24] integrates a tensor factorization and a deep neural network for drug  
46 synergy prediction. Models like DeepDDS [25] and DDoS [26] utilize graph neural networks over  
47 the molecular graph to enrich drug encoding. Further, Jiang's [27] and Hu's methods [28] expand  
48 the range of modalities employed for drug synergy predictions, including drug-drug and drug-target  
49 interactions. SDCNet [29] introduces the concept of cell line-specific graph representations for drug  
50 synergy data and trains a relational graph convolutional network over it.

51 Considering that drugs interact in the context of cell line treatment, we formulate the problem as a  
52 conditional variation of drug pair scoring framework [30] and call the cell line as a context of drug  
53 interaction. Further, we refer to it as "context" for simplification.

54 While existing approaches have demonstrated that sharing information between multiple modalities  
55 (fusion) leads to a performance gain [31], the fusion strategy in drug synergy is mostly a simple  
56 concatenation of latent representations, failing to capture the intrinsic dynamic synergies between  
57 drug pairs and cell lines.

58 Therefore, inspired by the concept of information fusion and the incorporation of a larger amount  
59 of contextual information in graph encoding, we introduce CongFu (Conditional Graph Fusion) for  
60 conditional drug pair scoring with a specific application of drug synergy prediction. The proposed  
61 layer includes context propagation and bottleneck, which work together to efficiently fuse two  
62 molecular graphs and a cell line. We present a technique for utilizing the proposed layer and  
63 evaluate the framework's performance on 12 benchmarks. The results indicate that our architecture  
64 outperforms existing approaches, and the inclusion of the CongFu layer tends to benefit other graph-  
65 based architectures. Additionally, we conduct ablation studies to emphasize the importance of every  
66 component in the proposed layer. Our explainability framework helps interpret model predictions,  
67 revealing the impact of drugs on specific genes.

68 To sum up, our contribution can be stated as follows: 1) We propose a novel CongFu layer for  
69 conditional graph pair scoring and apply it to drug synergy predictions 2) We conduct an ablation  
70 study to highlight the importance of fusion between graphs and to explore an appropriate place for  
71 initiating information sharing 3) We set the new state-of-the-art for 11 benchmarks derived from the  
72 DrugComb database in inductive and transductive setups 4) We provide the interpretability of our  
73 model to gain biological insights on gene-drug interactions.

## 74 2 Related Work

75 The related works can be categorized as follows:

76 **Linear Models.** Models such as Deep Synergy [19], MatchMaker [20], and AuDNNsynergy [22] uti-  
77 lize fully connected networks to process cell lines and drug features encoded via Morgan fingerprints.  
78 Deep Synergy applies a single MLP over the concatenated input triplet, while MatchMaker uses one  
79 MLP with shared weights to encode each drug conditionally based on the cell line. The hidden repre-  
80 sentations of the drug pairs are then passed to the MLP. AuDNNsynergy has a similar architecture  
81 to Deep Synergy but additionally processes "copy number variation" data, gene expressions, and  
82 mutations via autoencoders.

83 **Graph-based methods.** Models like DeepDDS [25] and DDoS [26] employ Message Passing  
84 Neural Networks (MPNNs) to encode each graph separately and an MLP to encode the cell line.  
85 All processed modalities are then concatenated and passed to the MLP. In models like SDCNet  
86 [29] and Jiang's method [27], the problem of drug synergy is formulated as link prediction. Both  
87 methods create cell line-specific heterogeneous networks of drugs and utilize an encoder-decoder  
88 architecture. Additionally, Jiang's method incorporates proteins into the drug-drug network, while Hu  
89 et al. [28] construct a single heterogeneous network of cell lines, drugs, diseases, and proteins. They

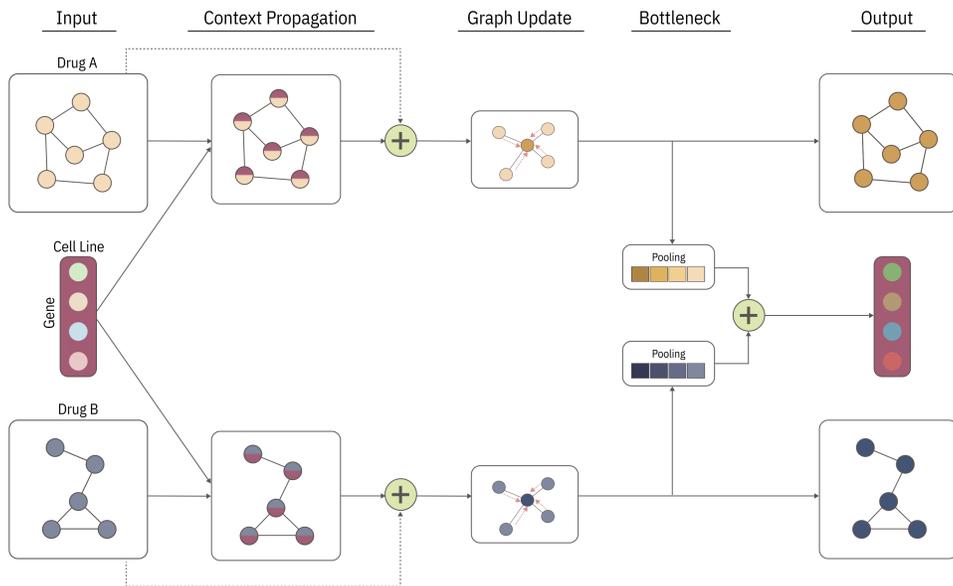


Figure 1: CongFu layer architecture. The layer takes in two graphs and a context vector as its input and produces an updated triplet as the output. This layer consists of three separate modules: Context Propagation, which facilitates the transfer of information from the initial context to the nodes of the graphs; Graph Update, which propagates the injected information throughout the graphs; and a Bottleneck that merges the local contexts related to each graph into a global context. The resulting output comprises updated representations of the two graphs and the context.

90 use the RotatE model [32] to encode diseases and pre-trained models [33, 34] for other modalities of  
 91 encoding. After propagation in the heterogeneous graph, drug and cell line embeddings are passed to  
 92 the MLP for the final prediction.

93 **Our method: CongFu.** Our proposed method, CongFu, is a significant advancement in the graph-  
 94 based category. We introduce a novel layer, CongFu, which formulates the heterogeneous graphs  
 95 between multiple drugs. This layer models dynamic interactions in a nonlinear manner, representing  
 96 a substantial improvement over existing methods. Additionally, we investigate a variety of strategies  
 97 to optimally integrate the CongFu layer, with the aim of maximizing predictive performance.

### 98 3 Methods

99 This section formalizes the CongFu layer and its components (Fig. 1). We start with the problem  
 100 formulation including notations and task descriptions. We then break down the architecture of the  
 101 CongFu layer into corresponding equations. Finally, we discuss its modularity characteristic, explore  
 102 potential use cases, and consider possible adaptation of the algorithm to support problems with  
 103 multiple (exceeding 2) graphs.

#### 104 3.1 Problem Formulation

105 **Notations.** Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  denote a graph with  $N$  nodes and  $E$  edges. Graph  $\mathcal{G}$  is associated with  
 106 an adjacency matrix  $\mathbf{A} \in \mathbb{R}^{N \times N}$ , where  $\mathbf{A}_{ij} = 0$  if there is no edge between nodes  $i$  and  $j$ ; node  
 107 feature set is denoted as  $\mathbf{X} \in \mathbb{R}^{N \times D_{node}}$ , while edge feature set is denoted as  $\mathbf{E} \in \mathbb{R}^{E \times D_{edge}}$

108 **Task: Conditional drug pair scoring.** Given a set of graph pairs  $\mathcal{D} = \{(\mathcal{G}_A^1, \mathcal{G}_B^1), \dots, (\mathcal{G}_A^n, \mathcal{G}_B^n)\}$   
 109 and associated context features for each graph pair  $\mathcal{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_n\}$  as input, where  $\mathbf{C}_i \in \mathbb{R}^{D_{cont}}$ ,  
 110 the objective of the task is to predict the corresponding target values  $\mathbb{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ , where  $\mathbf{Y}_i$   
 111 is a scalar value.

112 **3.2 CongFu Layer Architecture**

113 The CongFu layer receives as input two molecular graphs and a context vector (representing a cell  
114 line) and outputs an updated triplet. This layer consists of three distinct modules:

- 115 1. Context Propagation: allows the flow of information from the initial context to the nodes of  
116 the graphs
- 117 2. Graph Update: propagates the injected information along the graph
- 118 3. Bottleneck: combines the local contexts associated with each graph to form a global context

119 As a result, the layer outputs updated representations of both graphs and the context.

120 The context features  $\mathbf{C} \in \mathbb{R}^{1 \times D_{cont}}$  are required to have the same dimension ( $D_{cont}$ ) as node  
121 features  $\mathbf{X}_A \in \mathbb{R}^{N_A \times D_{node}}$ ,  $\mathbf{X}_B \in \mathbb{R}^{N_B \times D_{node}}$  ( $D_{node}$ ). To achieve this, the context features are  
122 linearly transformed as illustrated in (1), where  $\mathbf{W} \in \mathbb{R}^{D_{cont} \times D_{node}}$ . If context and node feature  
123 dimensions are equal ( $D_{cont} = D_{node}$ ), no transformation is required. Further, node and context  
124 feature dimensions are denoted as  $D$ .

$$\mathbf{C} = \mathbf{C}\mathbf{W} \tag{1}$$

125 The modular structure of the CongFu layer allows for the free choice of an aggregation strategy of  
126 initial and updated feature sets in (2), an MPNN for the *Graph Update* module (3), and a replacement  
127 of readout in the Bottleneck module (6).

128 **3.2.1 Context propagation**

129 This module updates node feature representations  $\mathbf{X}_A$  and  $\mathbf{X}_B$  based on the context  $\mathbf{C}$  using a  
130 conditional approach. The updated node representations  $\hat{\mathbf{X}}_A, \hat{\mathbf{X}}_B$  are then added to the initial node  
131 representations  $\mathbf{X}_A, \mathbf{X}_B$  (2). Specifically, we can express this process as follows:

$$\hat{\mathbf{X}}_j = \mathbf{X}_j + \mathbf{W}_1 \mathbf{X}_j + \mathbf{W}_2 \text{ReLU}(\mathbf{C}\mathbf{W}_3) \tag{2}$$

132 Here,  $j \in \{A, B\}$ ,  $\{\mathbf{X}_A, \mathbf{X}_B, \hat{\mathbf{X}}_A, \hat{\mathbf{X}}_B\} \in \mathbb{R}^{N \times D}$ , while  $\{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3\} \in \mathbb{R}^{D \times D}$ .

133 **3.2.2 Graph update**

134 After injecting the context into the node features, information is propagated along the graph. Due to  
135 the module’s modularity property, any MPNN (e.g., GIN [35], GraphSAGE [36], GPS [37]) can be  
136 used for graph updates (3). The edge features  $\mathbf{E}$  are optional and can be passed if available, and the  
137 MPNN supports them. Then, Batch Normalization [38] and ReLU [39] are applied over  $\hat{\mathbf{X}}_j$ , where  
138  $j \in A, B$ .

$$\hat{\mathbf{X}}_j = \text{MPNN}(\hat{\mathbf{X}}_j, \mathbf{E}_j, \mathbf{A}_j) \tag{3}$$

$$\hat{\mathbf{X}}_j = \text{BatchNorm}(\hat{\mathbf{X}}_j) \tag{4}$$

$$\hat{\mathbf{X}}_j = \text{ReLU}(\hat{\mathbf{X}}_j) \tag{5}$$

139 **3.2.3 Bottleneck**

140 In this step, the attention-based readout function  $\mathcal{F}$  is used to aggregate node-level features  $\hat{\mathbf{X}}_A, \hat{\mathbf{X}}_B$   
141 conditioned on the context  $\mathbf{C}$  (6). The resulting output,  $\mathbf{C}_A, \mathbf{C}_B$ , represents the local context of each  
142 graph.

$$\mathcal{F}(\hat{\mathbf{X}}_j; \mathbf{C}) \rightarrow \mathbf{C}_j \tag{6}$$

143 The equation of the readout function  $\mathcal{F}$ , inspired by the Graph Attention Networks (GAT) [40] layer  
 144 update rule, is illustrated in (7), where  $j \in A, B$ ,  $\mathbf{W} \in \mathbb{R}^{D \times D}$ ,  $d, b \in \mathbb{R}^D$  are trainable  
 145 matrix and biases respectively. The attention coefficient  $\alpha_i$  is computed in (8) to account for the  
 146 importance of context and node features.

$$\mathbf{C}_j = \sum_i^{N_j} \frac{1}{N_j + 1} \alpha_i \mathbf{W} \mathbf{X}_j^i + b \quad (7)$$

$$\alpha_i = \text{softmax}(d^T [\mathbf{W} \mathbf{C} \parallel \mathbf{W} \mathbf{X}_j^i]) \quad (8)$$

147 Finally, we update the global context  $\hat{\mathbf{C}}$  as a sum of local contexts  $\mathbf{C}_A$  and  $\mathbf{C}_B$  (9).

$$\hat{\mathbf{C}} = \mathbf{C}_A + \mathbf{C}_B \quad (9)$$

### 148 3.2.4 Extending to multiple graphs

149 While our algorithm is designed to operate over two graphs, it can be easily extended for multiple  
 150 (more than 2) graphs. The *Context Propagation* and *Graph Update* modules are invariant to the  
 151 number of graphs as each receives a single graph as input. The fusion process remains consistent  
 152 within the *Bottleneck*. Consequently, the sum over two graphs in (9) is replaced by the sum over  $k$   
 153 graphs' contexts.

$$\hat{\mathbf{C}} = \sum_j^k \mathbf{C}_j \quad (10)$$

## 154 4 Experiments

### 155 4.1 Datasets

156 We perform an extensive evaluation of the proposed algorithm for predicting drug synergy effects  
 157 using four datasets tested in three different setups, resulting in a total of 12 benchmarks. These  
 158 datasets are sourced from DrugComb [41], the most comprehensive and current database of drug  
 159 combinations. The goal is to predict the synergistic effect of drugs for a specific cell line (referred  
 160 to as the context in our notation), with the drugs represented as SMILES [42]. More details on the  
 161 dataset can be found in Appendix A.

### 162 4.2 Experimental Setup

163 We assess the model using three distinct setups for each dataset, adhering to the methodology proposed  
 164 by DeepDDS [25]. Specifically, we employ a transductive setup with a 5-fold cross-validation, where  
 165 the training set is further divided into training and validation subsets using a 90/10 ratio. In the  
 166 leave-drug-out setup, we partition the set of drugs into five equally sized groups, with the training set  
 167 excluding all drugs from the test set. We then conduct cross-validation stratified by drug groups. For  
 168 the leave-combination-out setup, the drug pairs from the test set are removed from the training set,  
 169 although individual drugs may still appear in both the training and test sets. The performance of the  
 170 model is evaluated using AUROC and AUPRC metrics to deal with imbalance.

### 171 4.3 Implementation details

172 The atom representation is computed as the embedded atomic number, while the edge representation  
 173 corresponds to the embedded bond type. The cell line is compressed into the latent space via a 2-layer  
 174 MLP. Each graph is individually encoded using a 3-layer Graph Isomorphism Network (GINE) [43]  
 175 with Batch Normalization [38] and ReLU activation. Two CongFu layers with the same embedding  
 176 dimension are applied to integrate information from the graphs and the cell line, utilizing GINE as  
 177 an MPNN. A prediction head consisting of a 2-layer MLP operates over the concatenation of the

Table 1: Comparison to SOTA on DrugComb - HSA Synergy Score

Method	Transductive		Leave-comb-out		Leave-drug-out	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
CongFu (ours)	<b>0.976 ± 0.001</b>	<b>0.949 ± 0.002</b>	<b>0.968 ± 0.003</b>	<b>0.931 ± 0.007</b>	<b>0.832 ± 0.02</b>	<b>0.67 ± 0.03</b>
DeepDDS	0.956 ± 0.02	0.913 ± 0.005	0.937 ± 0.008	0.87 ± 0.018	0.798 ± 0.03	0.625 ± 0.04
DSN - DDI	0.931 ± 0.01	0.861 ± 0.02	0.948 ± 0.004	0.894 ± 0.006	0.8 ± 0.013	0.645 ± 0.031
XGBoost	0.73 ± 0.005	0.565 ± 0.008	0.729 ± 0.005	0.556 ± 0.003	0.684 ± 0.022	0.48 ± 0.045
LogReg	0.723 ± 0.005	0.536 ± 0.007	0.718 ± 0.006	0.528 ± 0.013	0.67 ± 0.018	0.435 ± 0.043

Table 2: Ablation study on a conditional fusion on DrugComb - HSA Synergy Score

Method	Transductive		Leave-comb-out		Leave-drug-out	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
CongFu (ours)	<b>0.976 ± 0.001</b>	<b>0.949 ± 0.002</b>	<b>0.968 ± 0.003</b>	<b>0.931 ± 0.007</b>	<b>0.832 ± 0.02</b>	<b>0.67 ± 0.03</b>
w/o conditioning	0.966 ± 0.001	0.928 ± 0.003	0.949 ± 0.002	0.892 ± 0.007	0.812 ± 0.018	0.633 ± 0.04
w/o fusion	0.97 ± 0.017	0.939 ± 0.03	0.955 ± 0.002	0.906 ± 0.006	0.81 ± 0.02	0.607 ± 0.04

178 cell line and drug representations. ReLU is used as an activation function between hidden layers  
 179 in all MLPs. The training was conducted on a single NVIDIA RTX 6000 taking approximately 2  
 180 minutes per epoch. We utilized the Adam [44] optimizer and binary cross-entropy loss during training.  
 181 The training setup and hyperparameters remained consistent across all benchmarks, as illustrated in  
 182 Appendix D. The overall architecture is depicted in Appendix F.

## 183 4.4 Results

184 Our model, based on the CongFu layer, demonstrates superior performance compared to the other  
 185 methods in 11 out of the 12 benchmarks, according to the AUPRC score. The only exception is  
 186 observed in the Loewe leave-combination-out setup, where DeepDDS [25] marginally surpasses our  
 187 model by 0.003 in terms of AUROC. However, it’s worth noting that our model still outperforms  
 188 DeepDDS by 0.02 in AUPRC for the same setup. The most significant performance gap in favor  
 189 of our model is observed in the benchmarks using the HSA score (Table 1), with an improvement  
 190 ranging from 0.036 to 0.061 in AUPRC. Due to the large input vector of size 1508, XGBoost and  
 191 Logistic Regression struggle to capture all dependencies and show poor performance compared to  
 192 the state-of-the-art models. Tables 1, and other quantitative results in Appendix E, summarize our  
 193 results and compare CongFu-based architecture to other models mentioned in the experimental setup  
 194 section. Importantly, our model exhibits substantial improvement in inductive settings, underscoring  
 195 its capacity to generalize effectively to unseen data.

### 196 4.4.1 Ablation study

197 In order to validate the significance of conditional fusion, we perform an ablation study, focusing  
 198 on the transductive and leave-drug-out HSA benchmarks. The study is comprised of two main  
 199 experiments:

- 200 • *Without conditioning*: The aim of this experiment is to assess the impact of conditioning  
 201 on the context in the fusion component. We substitute the *Context Propagation* module  
 202 in CongFu with a cross-attention module. In this module, all nodes of  $\mathcal{G}_A$  and  $\mathcal{G}_B$  are  
 203 interconnected through a bipartite graph, and information exchange is facilitated via a GAT  
 204 as implemented in the intra-view of DSN-DDI. While this approach enables information  
 205 sharing between the two graphs, it does not account for the context between them.
- 206 • *Without fusion*: This experiment is designed to evaluate the importance of information  
 207 sharing (fusion) between graphs. We replace all CongFu layers with MPNNs - specifically,  
 208 GINE - to encode drugs independently. The results of this ablation study highlight the  
 209 significance of information exchange between the two graphs.

### 210 4.4.2 Determining the Optimal Point for Fusion

211 We conduct a series of experiments using a 5-layer model composed of MPNNs and fusion layers  
 212 (either CongFu or cross-attention layers). The fusion layers are applied after the MPNNs, which

213 initially encode each graph independently, without any information exchange. The idea behind  
 214 the middle fusion is, firstly, to let the model learn representations of drugs separately, then learn  
 215 the interaction of each drug with the cell line (as it was in MatchMaker), and finally combine  
 216 representations of drugs and a cell line together. The aim of this study is to highlight the importance  
 217 of conditional fusion (CongFu) between graphs and to identify the optimal point for initiating  
 218 information sharing, referred to as 'fusion injection'.

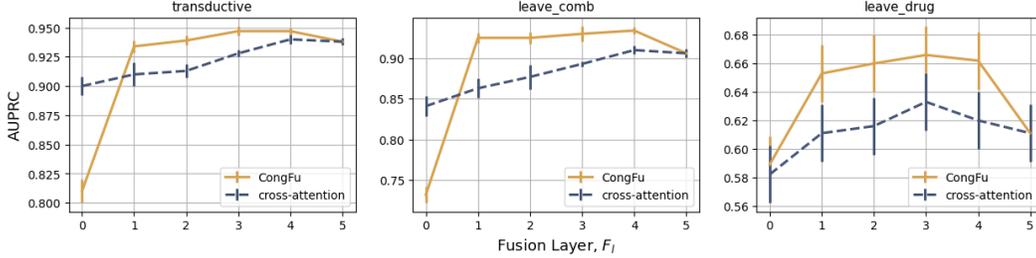


Figure 2: The impact of using the conditional fusion starting from different model layers on the leave-drug-out HSA benchmark

219 On Figure 2, the x-axis (Fusion layer) represents the layer in the model where fusion begins, i.e.,  
 220 the number of preceding MPNN layers. For instance, for  $F_l = 2$ , the model consists of 2 MPNN  
 221 layers followed by 3 fusion layers. Overall, starting from  $F_l = 1$ , the models with CongFu layers  
 222 consistently outperform those with cross-attention layers. The study reveals that in transductive and  
 223 leave-drug-out setups, the optimal model configuration includes 3 MPNN layers and 2 CongFu layers.  
 224 However, for the leave-comb setup, the model requires 4 MPNNs.

#### 225 4.4.3 Explainability

226 In this chapter, we provide biological insights by elucidating the predictions of the model. Our aim is  
 227 to answer the question: “What impact does each drug have on a specific gene?” To tackle this, we use  
 228 the chain rule to estimate the gradient magnitude of the output passed through drug encoders w.r.t a  
 229 specific gene.

230 Firstly, we calculate gradients of pooled drug embeddings from the last CongFu layer w.r.t a gene.  
 231 The gene is represented as  $g \in \mathbb{R}$ , the cell line is denoted as  $C = [g_1, g_2 \dots g_n] \in \mathbb{R}^N$ , and drug  
 232 embedding is denoted as  $h \in \mathbb{R}^D$ .

$$233 \mathcal{R}(g_i, h) = [\mathcal{R}(g_i, h_1), \mathcal{R}(g_i, h_2), \dots, \mathcal{R}(g_i, h_d)] = \left[ \frac{\partial h_1}{\partial g_i}, \frac{\partial h_2}{\partial g_i}, \dots, \frac{\partial h_n}{\partial g_i} \right] \quad (11)$$

233 Next, we compute gradients of the model output from the predictive head, denoted as  $y$ , w.r.t the drug  
 234 embeddings obtained from the last CongFu layer.

$$235 \mathcal{R}(h, y) = [\mathcal{R}(h_1, y), \mathcal{R}(h_2, y), \dots, \mathcal{R}(h_d, y)] = \left[ \frac{\partial y}{\partial h_1}, \frac{\partial y}{\partial h_2}, \dots, \frac{\partial y}{\partial h_d} \right] \quad (12)$$

235 Finally, we calculate the modulus of the dot product between these gradients multiplied by the input  
 236 value (gene), which represents the magnitude of the gradient passed through the drug encoder.

$$237 \mathcal{R}(g_i) = |g_i \mathcal{R}(g_i, h) \mathcal{R}(h, y)| = \left| g_i \sum_j \frac{\partial h_j}{\partial g_i} \frac{\partial y}{\partial h_j} \right| \quad (13)$$

237 To assess the impact of each drug on a gene, we compute the proportion of the magnitudes of drug A  
 238 and drug B.

239 Prior research [45, 46, 47] has demonstrated that the combination of the epidermal growth factor  
 240 receptor (EGFR) inhibitor Afatinib and the serine/threonine protein kinase B (AKT) inhibitor MK2206  
 241 has a synergistic impact on the treatment of lung cancer and head and neck squamous cell carcinoma.

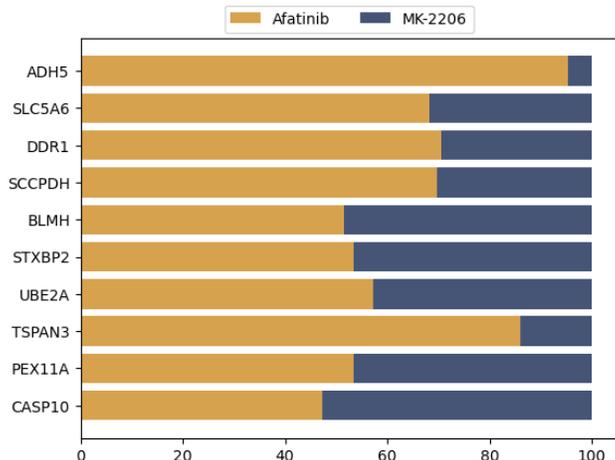


Figure 3: Impact of Afatinib and MK2206 top most important genes from BT-20 cell line

242 In DeepDDS [25], authors plot heat maps of the atom correlation matrix before and after training to  
 243 observe the change in feature patterns.

244 Figure 3 illustrates the impact of Afatinib and MK2206 on the top 10 most important genes from  
 245 BT-20 (breast tumor) cell line, sorted in descending order according to the magnitude of gradients.  
 246 Interestingly, these drugs have different impacts on each gene. For example, genes ADH5 and  
 247 TSPAN3 primarily interact with Afatinib rather than MK-2206. Though both drugs have quite similar  
 248 impacts on UBE2A and PEX11A genes. We believe the provided explainability framework will offer  
 249 scientists interesting biological insights regarding model prediction and will expedite the discovery of  
 250 new drugs.

## 251 5 Conclusion

252 In this work, we introduced a novel Conditional Graph Fusion Layer (CongFu) specifically designed  
 253 for drug synergy predictions. The CongFu layer utilizes an attention readout mechanism and a  
 254 bottleneck module to extract local graph contexts and conditionally fuse graph data within a global  
 255 context. The modular design of CongFu allows for easy customization by replacing layer modules,  
 256 such as readouts and graph encoders.

257 We conducted extensive experiments on four distinct datasets across three different setups to evaluate  
 258 CongFu’s performance in predicting drug synergy. CongFu outperformed state-of-the-art methods  
 259 on 11 out of 12 benchmark datasets, demonstrating its ability to capture complex drug synergy  
 260 patterns. Ablation studies further confirmed the importance of incorporating CongFu layers and their  
 261 contribution to the overall predictive performance.

262 By effectively predicting drug synergy in untested drug pairs, CongFu paves the way for optimizing  
 263 drug combinations and advancing personalized medicine. However, our study is not without limita-  
 264 tions. While we have developed a universal technique for solving the conditional graph pair scoring  
 265 problem, it is currently only applicable in the domain of drug synergy prediction. As new problems  
 266 that align with the task’s requirements emerge, our methodology can be applied and tested on them.  
 267 Although it is theoretically possible to extend the approach to a broader range of input graphs, the  
 268 lack of appropriate datasets prevents us from evaluating CongFu’s performance on these problems.  
 269 To the best of our knowledge, language models (LMs) have not been applied yet for predicting drug  
 270 synergy. We will consider combining LMs with the idea of fusion in future work.

271 **References**

- 272 [1] Reza Bayat Mokhtari, Tina S Homayouni, Narges Baluch, Evgeniya Morgatskaya, Sushil Kumar, Bikul  
273 Das, and Herman Yeger. Combination therapy in combating cancer. *Oncotarget*, 8(23):38022–38043, June  
274 2017.
- 275 [2] Shan Zhao, Tomohiro Nishimura, Yibang Chen, Evren U Azeloglu, Omri Gottesman, Chiara Giannarelli,  
276 Mohammad U Zafar, Ludovic Benard, Juan J Badimon, Roger J Hajjar, Joseph Goldfarb, and Ravi  
277 Iyengar. Systems pharmacology of adverse event mitigation by drug combinations. *Sci. Transl. Med.*,  
278 5(206):206ra140, October 2013.
- 279 [3] Jennifer O’Neil, Yair Benita, Igor Feldman, Melissa Chenard, Brian Roberts, Yaping Liu, Jing Li, Astrid  
280 Kral, Serguei Lejnine, Andrey Loboda, William Arthur, Razvan Cristescu, Brian B Haines, Christopher  
281 Winter, Theresa Zhang, Andrew Bloecher, and Stuart D Shumway. An unbiased oncology compound  
282 screen to identify novel combination strategies. *Mol. Cancer Ther.*, 15(6):1155–1162, June 2016.
- 283 [4] Jessica A Hill, Ron Ammar, Dax Torti, Corey Nislow, and Leah E Cowen. Genetic and genomic architecture  
284 of the evolution of resistance to antifungal drug combinations. *PLoS Genet.*, 9(4):e1003390, April 2013.
- 285 [5] Yejin Kim, Shuyu Zheng, Jing Tang, Wenjin Jim Zheng, Zhao Li, and Xiaoqian Jiang. Anticancer drug  
286 synergy prediction in understudied tissues using transfer learning. *Journal of the American Medical  
287 Informatics Association*, 28(1):42–51, 10 2020.
- 288 [6] Bissan Al-Lazikani, Udai Banerji, and Paul Workman. Combinatorial drug therapy for cancer in the  
289 post-genomic era. *Nat. Biotechnol.*, 30(7):679–692, July 2012.
- 290 [7] Erik De Clercq. The design of drugs for HIV and HCV. *Nat. Rev. Drug Discov.*, 6(12):1001–1018,  
291 December 2007.
- 292 [8] José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. Drug discovery with explainable artificial  
293 intelligence. *Nature Machine Intelligence*, 2(10):573–584, October 2020.
- 294 [9] Ning Yin, Wenzhe Ma, Jianfeng Pei, Qi Ouyang, Chao Tang, and Luhua Lai. Synergistic and antagonistic  
295 drug combinations depend on network topology. *PLoS One*, 9(4):e93960, 2014.
- 296 [10] Daphne Day and Lillian L Siu. Approaches to modernize the combination drug development paradigm.  
297 *Genome Med.*, 8(1):115, October 2016.
- 298 [11] Kaifang Pang, Ying-Wooi Wan, William T. Choi, Lawrence A. Donehower, Jingchun Sun, Dhruv Pant, and  
299 Zhandong Liu. Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics*,  
300 30(10):1456–1463, 01 2014.
- 301 [12] Peng Li, Chao Huang, Yingxue Fu, Jinan Wang, Ziyin Wu, Jinlong Ru, Chunli Zheng, Zihu Guo, Xuetong  
302 Chen, Wei Zhou, Wenjuan Zhang, Yan Li, Jianxin Chen, Aiping Lu, and Yonghua Wang. Large-scale  
303 exploration and analysis of drug combinations. *Bioinformatics*, 31(12):2007–2016, June 2015.
- 304 [13] Ricardo Macarron, Martyn N Banks, Dejan Bojanic, David J Burns, Dragan A Cirovic, Tina Garyantes,  
305 Darren V S Green, Robert P Hertzberg, William P Janzen, Jeff W Paslay, Ulrich Schopfer, and G Sitta  
306 Sittampalam. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.*,  
307 10(3):188–195, March 2011.
- 308 [14] Nikko P Torres, Anna Y Lee, Guri Giaever, Corey Nislow, and Grant W Brown. A high-throughput yeast  
309 assay identifies synergistic drug combinations. *Assay Drug Dev. Technol.*, 11(5):299–307, June 2013.
- 310 [15] Liye He, Evgeny Kuleskiy, Jani Saarela, Laura Turunen, Krister Wennerberg, Tero Aittokallio, and Jing  
311 Tang. Methods for high-throughput drug combination screening and synergy scoring. *Methods Mol. Biol.*,  
312 1711:351–398, 2018.
- 313 [16] Xing Chen, Biao Ren, Ming Chen, Ming-Xi Liu, Wei Ren, Quan-Xin Wang, Li-Xin Zhang, and Gui-Ying  
314 Yan. Asdcd: Antifungal synergistic drug combination database. *PLOS ONE*, 9(1):1–6, 01 2014.
- 315 [17] Susan L Holbeck, Richard Camalier, James A Crowell, Jeevan Prasaad Govindharajulu, Melinda Holling-  
316 shead, Lawrence W Anderson, Eric Polley, Larry Rubinstein, Apurva Srivastava, Deborah Wilsker, Jerry M  
317 Collins, and James H Doroshow. The national cancer institute ALMANAC: A comprehensive screen-  
318 ing resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res.*,  
319 77(13):3564–3576, July 2017.
- 320 [18] Nitin Dogra and Vijay Chahar. A comprehensive review on deep synergistic drug prediction techniques for  
321 cancer. *Archives of Computational Methods in Engineering*, 06 2021.

- 322 [19] Kristina Preuer, Richard P I Lewis, Sepp Hochreiter, Andreas Bender, Krishna C Bulusu, and Günter  
323 Klambauer. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*,  
324 34(9):1538–1546, 12 2017.
- 325 [20] Halil Ibrahim Kuru, Ozgur Tastan, and A. Ercument Cicek. Matchmaker: A deep learning framework for  
326 drug synergy prediction. *bioRxiv*, 2020.
- 327 [21] H. L. Morgan. The generation of a unique machine description for chemical structures—a technique  
328 developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 1965.
- 329 [22] Tianyu Zhang, Liwei Zhang, Philip R. O. Payne, and Fuhai Li. Synergistic drug combination prediction by  
330 integrating multi-omics data in deep learning models, 2018.
- 331 [23] Qiao Liu and Lei Xie. Transynergy: Mechanism-driven interpretable deep neural network for the synergistic  
332 prediction and pathway deconvolution of drug combinations. *PLOS Computational Biology*, 17(2):1–22,  
333 02 2021.
- 334 [24] Zexuan Sun, Shujun Huang, Peiran Jiang, and Pingzhao Hu. DTF: Deep Tensor Factorization for predicting  
335 anticancer drug synergy. *Bioinformatics*, 36(16):4483–4489, 05 2020.
- 336 [25] Jinxian Wang, Xuejun Liu, Siyuan Shen, Lei Deng, and Hui Liu. DeepDDS: deep graph neural network  
337 with attention mechanism to predict synergistic drug combinations. *Briefings in Bioinformatics*, 23(1), 09  
338 2021. bbab390.
- 339 [26] Kyriakos Schwarz, Alicia Pliego-Mendieta, Lara Planas-Paz, Chantal Pauli, Ahmed Allam, and Michael  
340 Krauthammer. Ddos: A graph neural network based drug synergy prediction algorithm, 2022.
- 341 [27] Peiran Jiang, Shujun Huang, Zexuan Sun, Ted Lakowski, and Pingzhao Hu. Deep graph embedding  
342 for prioritizing synergistic anticancer drug combinations. *Computational and Structural Biotechnology  
343 Journal*, 18:427–438, 2020.
- 344 [28] Zhihang Hu, Qinze Yu, Yucheng Guo, Taifeng Wang, Irwin King, Xin Gao, Le Song, and Yu Li. Drug  
345 synergistic combinations predictions via large-scale pre-training and graph structure learning, 2023.
- 346 [29] Peng Zhang, Shikui Tu, Wen Zhang, and Lei Xu. Predicting cell line-specific synergistic drug combinations  
347 through a relational graph convolutional network with attention mechanism. *Briefings in Bioinformatics*,  
348 23(6), 09 2022. bbac403.
- 349 [30] Benedek Rozemberczki, Stephen Bonner, Andriy Nikolov, Michael Ughetto, Sebastian Nilsson, and Eliseo  
350 Papa. A unified view of relational deep learning for drug pair scoring. *arXiv preprint arXiv:2111.02916*,  
351 2021.
- 352 [31] Arsha Nagrani, Shan Yang, Anurag Arnab, Aren Jansen, Cordelia Schmid, and Chen Sun. Attention  
353 bottlenecks for multimodal fusion, 2022.
- 354 [32] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by  
355 relational rotation in complex space, 2019.
- 356 [33] Han Li, Dan Zhao, and Jianyang Zeng. KPGT. In *Proceedings of the 28th ACM SIGKDD Conference on  
357 Knowledge Discovery and Data Mining*. ACM, aug 2022.
- 358 [34] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Demi Guo, Myle Ott,  
359 C. Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling  
360 unsupervised learning to 250 million protein sequences. *bioRxiv*, 2020.
- 361 [35] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?,  
362 2019.
- 363 [36] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs,  
364 2018.
- 365 [37] Ladislav Rampásek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique  
366 Beaini. Recipe for a general, powerful, scalable graph transformer, 2023.
- 367 [38] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing  
368 internal covariate shift, 2015.
- 369 [39] Kunihiro Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*,  
370 20:121–136, 1975.

- 371 [40] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio.  
372 Graph attention networks, 2018.
- 373 [41] Shuyu Zheng, Jehad Aldahdooh, Tolou Shadbahr, Yinyin Wang, Dalal Aldahdooh, Jie Bao, Wenyu Wang,  
374 and Jing Tang. DrugComb update: a more comprehensive drug sensitivity data repository and analysis  
375 portal. *Nucleic Acids Res.*, 49(W1):W174–W184, July 2021.
- 376 [42] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology  
377 and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.
- 378 [43] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec.  
379 Strategies for pre-training graph neural networks, 2020.
- 380 [44] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- 381 [45] Helmout Modjtahedi, Byoung Chul Cho, Martin C Michel, and Flavio Solca. A comprehensive review of  
382 the preclinical efficacy profile of the ErbB family blocker afatinib in cancer. *Naunyn. Schmiedebergs. Arch.*  
383 *Pharmacol.*, 387(6):505–521, June 2014.
- 384 [46] Renato José Silva-Oliveira, Matias Melendez, Olga Martinho, Maicon F Zanon, Luciano de Souza Viana,  
385 André Lopes Carvalho, and Rui Manuel Reis. AKT can modulate the in vitro response of HNSCC cells to  
386 irreversible EGFR inhibitors. *Oncotarget*, 8(32):53288–53301, August 2017.
- 387 [47] Ming-Szu Hung, I-Chuan Chen, Jr-Hau Lung, Paul-Yann Lin, Ya-Chin Li, and Ying-Huang Tsai. Epidermal  
388 growth factor receptor mutation enhances expression of cadherin-5 in lung cancer cells. *PLoS One*,  
389 11(6):e0158395, June 2016.
- 390 [48] Greg Landrum. Rdkit: Open-source cheminformatics software. 2016.
- 391 [49] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR*  
392 *Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- 393 [50] Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu,  
394 Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, David L Lahr, Jodi E Hirschman, Zihan  
395 Liu, Melanie Donahue, Bina Julian, Mariya Khan, David Wadden, Ian C Smith, Daniel Lam, Arthur  
396 Liberzon, Courtney Toder, Mukta Bagul, Marek Orzechowski, Oana M Enache, Federica Piccioni, Sarah A  
397 Johnson, Nicholas J Lyons, Alice H Berger, Alykhan F Shamji, Angela N Brooks, Anita Vrcic, Corey Flynn,  
398 Jacqueline Rosains, David Y Takeda, Roger Hu, Desiree Davison, Justin Lamb, Kristin Ardlie, Larson  
399 Hogstrom, Peyton Greenside, Nathanael S Gray, Paul A Clemons, Serena Silver, Xiaoyun Wu, Wen-Ning  
400 Zhao, Willis Read-Button, Xiaohua Wu, Stephen J Haggarty, Lucienne V Ronco, Jesse S Boehm, Stuart L  
401 Schreiber, John G Doench, Joshua A Bittker, David E Root, Bang Wong, and Todd R Golub. A next  
402 generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171(6):1437–1452.e17,  
403 November 2017.
- 404 [51] Zimeng Li, Shichao Zhu, Bin Shao, Xiangxiang Zeng, Tong Wang, and Tie-Yan Liu. DSN-DDI: an  
405 accurate and generalized framework for drug–drug interaction prediction by dual-view representation  
406 learning. *Briefings in Bioinformatics*, 24(1), 01 2023. bbac597.
- 407 [52] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series*  
408 *B (Methodological)*, 20(2):215–232, 1958.
- 409 [53] Tianqi Chen and Carlos Guestrin. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International*  
410 *Conference on Knowledge Discovery and Data Mining*. ACM, aug 2016.
- 411 [54] W R Greco, G Bravo, and J C Parsons. The search for synergy: a critical review from a response surface  
412 perspective. *Pharmacol. Rev.*, 47(2):331–385, June 1995.
- 413 [55] C.I. Bliss. The toxicity of poisons applied jointly. *Annals of Applied Biology*, 26:585 – 615, 06 2008.
- 414 [56] Bhagwan Yadav, Krister Wennerberg, Tero Aittokallio, and Jing Tang. Searching for drug synergy in  
415 complex dose-response landscapes using an interaction potency model. *Comput. Struct. Biotechnol. J.*,  
416 13:504–513, September 2015.
- 417 [57] M C Berenbaum. What is synergy? *Pharmacological Reviews*, 41(2):93–141, 1989.

## 418 A Dataset details

419 Drugs represented as SMILES were converted using RDKit [48] into a PyG [49] graph, with atoms  
420 represented as nodes and bonds represented as edges. The features of the cell line are gathered from  
421 the Genomics of Drug Sensitivity in Cancer <sup>1</sup>. This includes normalized basal expression profiles of  
422 approximately 1000 human cancer cell lines. From the normalized expression levels of 17737 genes,  
423 we select 908 landmark genes [50]. Therefore, the cell lines are represented by a feature vector of  
424 length 908.

## 425 B Baselines

426 CongFu is benchmarked against state-of-the-art methods for drug synergy predictions. Specifically,  
427 we utilize the official implementations of DeepDDS [25] and a modified version of DSN-DDI [51]. In  
428 the latter, the relation-type embedding is substituted with a cell feature matrix, enabling the prediction  
429 of drug synergy from drug-drug interactions. For the baseline models, we implement Logistic  
430 Regression [52] and XGBoost [53] over the concatenated representations of cell lines and drugs,  
431 encoded using a pre-trained Deep Graph Infomax model [43]. SDCNet [29] and Hu’s methods [28]  
432 are not included in the comparison due to the irreproducibility and absence of codebase, respectively.

## 433 C Preprocessing description

434 Each dataset is created by quantifying the target through four distinct types of synergy scores,  
435 specifically Loewe additivity (Loewe) [54], Bliss independence (Bliss) [55], zero interaction potency  
436 (ZIP) [56], and highest single agent (HSA) [57]. These targets describe the measurement of drug  
437 interaction, specifically the degree of additional drug responses observed compared to the expected  
438 response. In other words, drug synergy indicates the percentage of excess or reduced response in  
439 antagonistic settings.

440 The preprocessing of the dataset follows the strategy of DDoS [26]. Initially, we exclude all triplets  
441 that do not have corresponding identifiers in the cell line feature table. Then, triplets with any missing  
442 data (cell line, drugs, targets) are filtered out. Finally, duplicated triplets are removed. Each of the  
443 four synergy scores is binarized based on thresholds [15]. Samples with a synergy score above 10 are  
444 considered positive (synergistic), and samples lower than -10 are considered negative (antagonistic).  
445 Consequently, we end up with four datasets (Loewe, Bliss, HSA, and ZIP) with names corresponding  
446 to their targets. The statistics of each dataset are described in Table 3.

Table 3: Statistics of the datasets, where Loewe, Bliss, HSA, ZIP - datasets derived from DrugComb. Statistics for DrugComb are calculated after the preprocessing stage. The percentage of the positive labels is rounded to the first decimal point.

Dataset	# Samples	% Positive Labels	# Drugs	# Cell Lines
Loewe	163816	14.8	2147	164
Bliss	125548	49.5	1868	164
HSA	108559	29.5	1189	162
ZIP	89047	59.8	1810	162
DrugComb	647232	-	4268	288

<sup>1</sup>[https://www.cancerrxgene.org/gdsc1000/GDSC1000\\_WebResources/Data/preprocessed/Cell\\_line\\_RMA\\_proc\\_basalExp.txt.zip](https://www.cancerrxgene.org/gdsc1000/GDSC1000_WebResources/Data/preprocessed/Cell_line_RMA_proc_basalExp.txt.zip)

447 **D Model hyperparameters**

Table 4: Model hyperparameters

Hyperparameter	Value	Hyperparameter	Value
Learning rate	$1e - 4$	Epochs	100
Node embedding size	300	Edge embedding size	300
# Graph encoders	3	Graph encoder	GINE([300, 300, 300])
# Graph encoders (CongFu)	2	Graph encoder (CongFu)	GINE([300, 300, 300])
Cell line encoder	[908, 512, 300]	Prediction Head	[812, 256, 64]

448 **E Additional experimental results**

Table 5: Comparison to SOTA on DrugComb - Bliss Synergy Score

Method	Transductive		Leave-comb-out		Leave-drug-out	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
CongFu (ours)	<b>0.982 ± 0.001</b>	<b>0.981 ± 0.001</b>	<b>0.975 ± 0.002</b>	<b>0.974 ± 0.003</b>	<b>0.79 ± 0.02</b>	<b>0.779 ± 0.02</b>
DeepDDS	0.956 ± 0.004	0.955 ± 0.004	0.941 ± 0.009	0.938 ± 0.009	0.76 ± 0.03	0.75 ± 0.03
DSN - DDI	0.894 ± 0.04	0.886 ± 0.04	0.952 ± 0.003	0.946 ± 0.004	0.754 ± 0.005	0.742 ± 0.009
XGBoost	0.717 ± 0.003	0.652 ± 0.004	0.712 ± 0.005	0.647 ± 0.011	0.64 ± 0.009	0.584 ± 0.012
LogReg	0.664 ± 0.003	0.605 ± 0.003	0.661 ± 0.004	0.603 ± 0.01	0.595 ± 0.011	0.552 ± 0.012

Table 6: Comparison to SOTA on DrugComb - Loewe Synergy Score

Method	Transductive		Leave-comb-out		Leave-drug-out	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
CongFu (ours)	<b>0.939 ± 0.005</b>	<b>0.791 ± 0.01</b>	0.772 ± 0.04	0.403 ± 0.07	0.774 ± 0.03	<b>0.423 ± 0.07</b>
DeepDDS	0.926 ± 0.004	0.746 ± 0.018	<b>0.775 ± 0.03</b>	<b>0.409 ± 0.06</b>	<b>0.777 ± 0.027</b>	0.403 ± 0.07
DSN - DDI	0.807 ± 0.02	0.437 ± 0.039	0.771 ± 0.028	0.358 ± 0.02	0.774 ± 0.019	0.361 ± 0.033
XGBoost	0.621 ± 0.002	0.303 ± 0.003	0.562 ± 0.013	0.215 ± 0.026	0.562 ± 0.013	0.215 ± 0.026
LogReg	0.61 ± 0.004	0.27 ± 0.005	0.58 ± 0.02	0.21 ± 0.028	0.58 ± 0.02	0.21 ± 0.028

Table 7: Comparison to SOTA on DrugComb - ZIP Synergy Score

Method	Transductive		Leave-comb-out		Leave-drug-out	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
CongFu (ours)	<b>0.986 ± 0.002</b>	<b>0.99 ± 0.001</b>	<b>0.983 ± 0.001</b>	<b>0.988 ± 0.001</b>	<b>0.829 ± 0.01</b>	<b>0.874 ± 0.01</b>
DeepDDS	0.977 ± 0.003	0.983 ± 0.002	0.964 ± 0.005	0.974 ± 0.004	0.812 ± 0.008	0.86 ± 0.01
DSN - DDI	0.947 ± 0.01	0.96 ± 0.009	0.964 ± 0.002	0.974 ± 0.002	0.793 ± 0.024	0.844 ± 0.012
XGBoost	0.736 ± 0.002	0.742 ± 0.001	0.732 ± 0.002	0.739 ± 0.005	0.665 ± 0.015	0.689 ± 0.027
LogReg	0.692 ± 0.003	0.712 ± 0.002	0.691 ± 0.008	0.711 ± 0.007	0.619 ± 0.008	0.66 ± 0.025

## 449 F CongFu schematics

### 450 F.1 CongFu-based model architecture

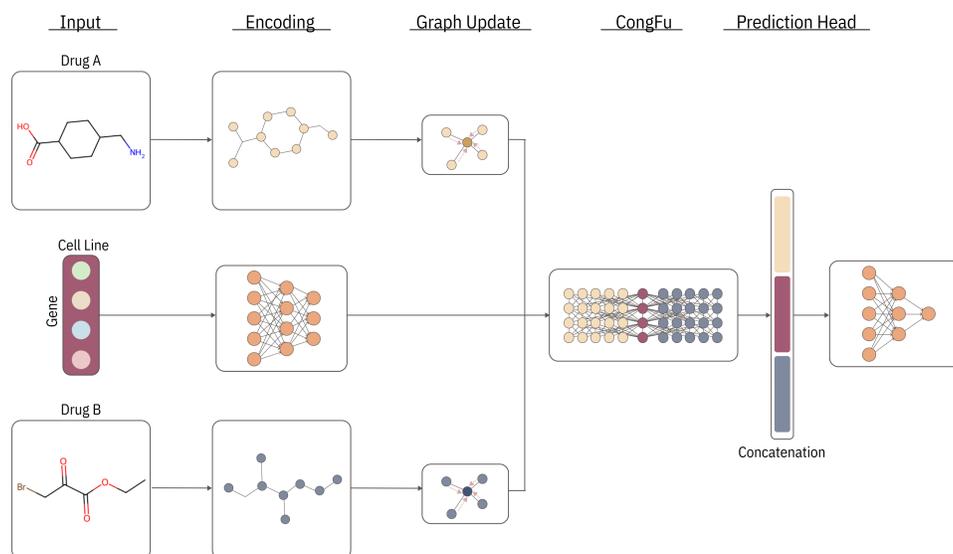


Figure 4: CongFu-based model architecture. The model takes in two drugs and a cell line as its input. The pairs of drugs are represented as graphs, where node features for each are obtained via atomic number embeddings, edge features are calculated as bond embeddings, and the cell line is encoded using MLP. Then, MPNN(s) are used to encode each graph separately. Next, CongFu layer(s) is utilized to fuse information from graphs and a cell line. Finally, an MLP is applied over the concatenation of drugs and the cell lines.

---

**Algorithm 1:** Algorithm for  $L$  layer CongFu-based model.

---

**Input:** Pair of graphs  $\mathcal{G}_A, \mathcal{G}_B$  with  $N_A, N_B$  nodes, and  $\mathbf{E}_A, \mathbf{E}_B$  edges; Adjacency matrices  $\mathbf{A}_A \in \mathbb{R}^{N_A \times N_A}$  and  $\mathbf{A}_B \in \mathbb{R}^{N_B \times N_B}$ ; Node features  $\mathbf{X}_A \in \mathbb{R}^{N_A \times 1}, \mathbf{X}_B \in \mathbb{R}^{N_B \times 1}$ , and edge features  $\mathbf{E}_A \in \mathbb{R}^{E_A \times 1}, \mathbf{E}_B \in \mathbb{R}^{E_B \times 1}$ ; Context  $\mathbf{C} \in \mathbb{R}^{1 \times D_{cont}}$ ; Layer  $l \in [0, L - 1]$ ; Fusion layer  $F_l$ .

**Output:** Node features  $\mathbf{X}_A \in \mathbb{R}^{N_A \times D}, \mathbf{X}_B \in \mathbb{R}^{N_B \times D}$ ; Context  $\mathbf{C} \in \mathbb{R}^{1 \times D}$ ;

$\mathbf{C} \leftarrow \text{MLP}(\mathbf{C}) \in \mathbb{R}^{1 \times D}$

**for**  $j \in [A, B]$  **do**

$\mathbf{X}_j \leftarrow \text{NodeEncoder}(\mathbf{X}_j) \in \mathbb{R}^{N_j \times D}$

**for**  $l = 0, 1, \dots, F_l - 1$  **do**

$\mathbf{E}_j \leftarrow \text{EdgeEncoder}(\mathbf{E}_j) \in \mathbb{R}^{E_j \times D}$

$\mathbf{X}_j \leftarrow \text{MPNN}_l(\mathbf{X}_j, \mathbf{E}_j, \mathbf{A}_j)$

$\mathbf{X}_j = \text{BatchNorm}(\mathbf{X}_j)$

$\mathbf{X}_j = \text{ReLU}(\mathbf{X}_j)$

**for**  $l = F_l, F_l + 1, \dots, L - 1$  **do**

**for**  $j \in [A, B]$  **do**

$\mathbf{X}_j \leftarrow \mathbf{X}_j + \mathbf{W}_1^l \mathbf{X}_j + \mathbf{W}_2^l \text{ReLU}(\mathbf{W}_3^l \mathbf{C})$

$\mathbf{X}_j \leftarrow \text{MPNN}_l(\mathbf{X}_j, \mathbf{E}_j, \mathbf{A}_j)$

$\mathbf{X}_j \leftarrow \text{BatchNorm}(\mathbf{X}_j)$

$\mathbf{X}_j \leftarrow \text{ReLU}(\mathbf{X}_j)$

**for**  $i = 0, 1, \dots, N_j$  **do**

$\alpha_i \leftarrow \text{softmax}(\mathbf{d}_i^T [\mathbf{W}_4^l \mathbf{C} \parallel \mathbf{W}_4^l \mathbf{X}_j^i])$

$\mathbf{C}_j \leftarrow \sum_i^{N_j} \frac{1}{N_j + 1} \alpha_i \mathbf{W}_4^l \mathbf{X}_j^i + \mathbf{b}_l$

$\mathbf{C} \leftarrow \mathbf{C}_A + \mathbf{C}_B$

**return**  $\text{MLP}([\mathbf{X}_A \parallel \mathbf{X}_B \parallel \mathbf{C}]) \in \mathbb{R}$

---