EasyChair Preprint

№ 14299

Distribution Matching for Drug-Drug Interaction Prediction

Olivia Lee and Su Yu

August 5, 2024

Distribution Matching for Drug-Drug Interaction Prediction

Cuiyu Li[1] and Yu Su[2]

[1] Zhejiang Laboratory, Hangzhou, Zhejiang 311100, China.
oliviluckyday@gmail.com
[2] Changchun University of Science and Technology suy410626735@gmail.com

Abstract. Drug-Drug interactions (DDIs) refer to the mutual effects that may occur when two or more drugs are co-administered. These interactions have significant implications for the efficacy, safety, or tolerability of medications, making them of paramount importance in the field of medicine. Recent research has focused on deep learning techniques, such as graph-based learning methods, which typically consider the molecular structure information of drugs but often neglect inter-view information.To overcome this limitation, we propose a hierarchical graph-based deep learning method that efficiently aligns intra-view and inter-view embeddings. Specifically, we perform distribution matching in both the feature space and output space, maximizing mutual information between the two views to enhance prediction accuracy. Additionally, we introduce a novel loss function that uses central matching distribution (CMD) to balance information between intra-view and inter-view embeddings, instead of relying on unsupervised contrastive learning. This approach increases computational speed by 50% while maintaining high accuracy. Evaluations on three datasets demonstrate that our method outperforms other state-of-the-art models in DDI prediction.

Keywords: hierarchical graph representation· central matching distribution· drug-drug interaction· graph embedding

1 Introduction

Combination therapies are frequently employed to manage multiple illnesses and diminish drug resistance [1]. However, administering multiple drugs concurrently can lead to Drug-Drug Interactions (DDIs), a prevalent source of medication errors. Notably, the elderly, with a prevalence ranging from 20-40%, are susceptible to such interactions [2]. As depicted in Figure 1, solid lines represent established drug-drug interactions, while dashed lines signify unknown or yet-to-be predicted drug-drug interaction. Detecting DDIs poses challenges, and conventional approaches are both expensive and time-intensive. Consequently, the field of molecular computational science is indispensable for enhancing DDI prediction.

Computational biology, primarily anchored in statistical methodologies, boasts a wide-ranging array of applications. It serves as a tool to forecast protein structures and unravel their functions and interactions [3]. In the field of genomics,

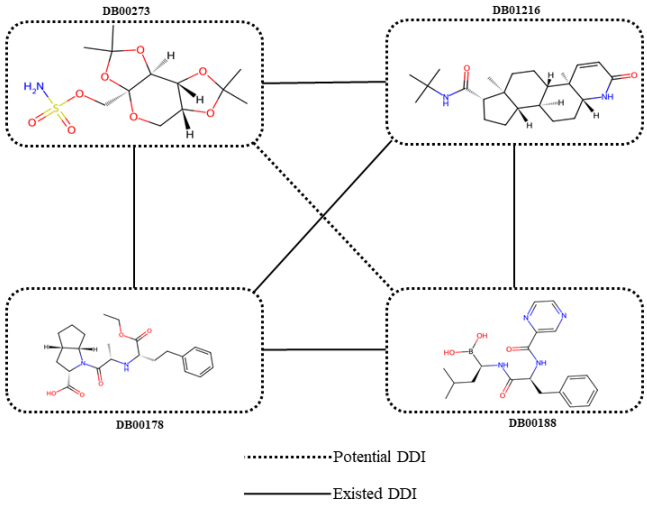

Fig. 1: DDI hierarchical graph representation

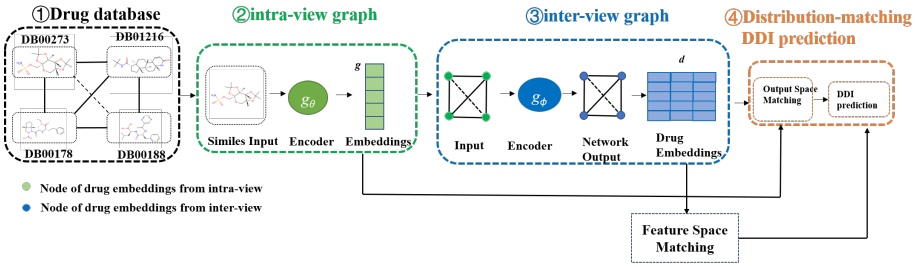

Fig. 2: The illustrative schematic diagram of our proposed **DM-DDI** framework. It comprises three consecutive and interrelated phases: 1) During the initial phase, drug molecular graph underdo encoding into drug embeddings using a message-passing network that is sensitive to bonds. Subsequently, attentive pooling is employed. 2) In the second phase, external DDI relationships are incorporated into embeddings through a GCN encoder. Additionally, the distribution matching of feature space matching and output space feature mapping is utilized to harmonize information from diverse perspectives in order to update the drug embeddings. 3) Lastly, leveraging the acquired drug embeddings, we have developed an interaction predictor responsible for producing the ultimate prediction outcomes.

it plays a pivotal role in gene identification, gene expression analysis, and the classification of DNA sequences[4–6]. Furthermore, computational biology finds utility in scrutinizing disease mutations [7] and facilitating the discovery of novel drugs[8, 9].

In the realm of Drug-Drug Interaction (DDI) detection, machine learning approaches have emerged as potent tools to enhance efficiency. Traditional machine learning methods have historically leaned on similarity-based features to predict DDIs, assuming that drugs with analogous chemical structures exhibit comparable DDI patterns. These approaches necessitate manual extraction of drug features and their subsequent incorporation into machine learning models for DDI prognosis [10, 11, 37, 13]. Nevertheless, these techniques are time-intensive and constrained by their reliance on manual feature engineering, leading to a gradual phasing-out of such methodologies.

Recently, there has been a surge in interest in graph-based research, largely due to the evolution of graph neural networks (GNNs). These graph-based approaches can be broadly categorized into two classes: Molecular graph-based methods and hierarchical graph-based methods. Molecular graph-based techniques rely exclusively on GNNs to extract structural features from drug molecules for predicting drug interactions [45, 30, 34, 17, 19]. Although these molecular graph-based methods have effectively mitigated the reliance on feature engineering and demonstrated promising results in DDI prediction tasks, they overlook the topological relationships between drugs. To overcome this limitation, several studies have delved into hierarchical models capable of generating intricate semantic representations and eliminating the need for manual feature engineering [20].

The utilization of hierarchical graph models has proven to be a potent approach for harnessing both molecular and topological information in the effective prediction of DDIs [44, 22, 23, 46, 25]. Figure 1 provides a visual representation of a DDI network comprising drug entities, where each node symbolizes an instance of a drug molecule, and each link signifies a verified interaction between two drugs. Furthermore, each drug node is portrayed as a molecular graph. To implement this hierarchical graph structure for DDI prediction, a molecular graph model is employed to encode drug molecules. The results obtained from these molecular graphs act as inputs to the drug interaction network graph, ultimately yielding the final representations of the drugs.

A key aspect in the hierarchical graph paradigm is how to maximize the mutual information between inter-view and intra-view information, which we refer to as distribution matching. The previous work [22] explored distribution matching in the output space, specifically by computing the Kullback–Leibler divergence (KLD) [27] between inter-view prediction probabilities and intra-view prediction probabilities to achieve distribution matching. However, they did not address matching in the feature space.

Building upon this foundation, the paper [28] employed contrastive learning to achieve distribution matching of inter-view features and intra-view features, effectively maximizing mutual information. The challenge here lies in the need to construct positive and negative samples using the topological relationships

of the upper-level graph, resulting in a computational complexity of $O(n^2)$. To address this, we propose Distribution Matching Drug-Drug Interaction prediction method, DM-DDI, a flexible and efficient feature space matching algorithm, leveraging O(n) methods such as CMD (central moment discrepancy) [29] to achieve both feature space distribution matching and output space distribution matching. Our experimental results demonstrate that our matching approach outperforms previous methods, achieving state-of-the-art results on two out of three datasets and comparable performance on the remaining dataset. In additionally, the nerual network architechture has achivied 1 50% increase in computational speed.

2 Methodology

In this section, we present a novel approach to predict Drug-Drug Interactions, namely DM-DDI (Distribution Matching for Drug-Drug Interaction). Next, we will elaborate on the details of the DM-DDI.

2.1 Datasets

We conducted assessments of our proposed model DM-DDI on three various scales of benchmark datasets, i.e., the small-scale dataset *ZhangDDI*[3], the medium-scale dataset *ChCh-Miner*[4] and the large-scale dataset *DeepDDI*[5]. The detailed statictics information of three datasets are shown in the Table1. Three datasets have their own attributes, such as *ZhangDDI* dataset comprises a relatively small number of drugs but contains completely fingerprints for each drugs. In contrast, the*DeepDDI* which is on a larger scale, exhibits a high prevalence of missing figerprints fo most drugs. As for the *ChCh-Miner* dataset, despite having nearly three times the number of drugs compared to the *ZhangDDI* dataset, it contains an equivlent number of labeled DDI links. In our experiments, we only used data items that can be converted into molecular graphs for learning from SMILES strings.

Table 1: Detailed information about the DDI datasets.

Dataset	Drugs	DDI links	Size	Information
ZhangDDI	548	48548	Small	Similarity
ChCh-Miner	1514	48514	Medium	/
DeepDDI	1861	192284	Large	Polypharmacy side-effect[**?**]

[3] https://github.com/zw9977129/drug-drug-interaction/tree/master/dataset
[4] http://snap.stanford.edu/biodata/datasets/10001/10001-ChCh-Miner.html
[5] https://zenodo.org/record/1205795

2.2 Preminaries

To establish the framework for our proposed model DM-DDI, we initially present the essential notations utilized in this paper. We represent the collection of drugs as $D = \{d_1, ..., d_N\}$, where N represents the total count of drugs. Each drug is depicted through a molecular graph deneted as G. These drugs, connected through established DDIs, collectively from the DDI network graph $\mathcal{G}$, which serves as the foundation for their interrelationships. The subsequent sections provide detailed descriptions of the molecular graph, DDI network graph, and the DDI problem.

Molecular Graph. The molecular graph which is denoted as $G = (A, B)$, where $A \in R^{n \times d_h}$ signifies the matrix containing atom-level characteristics, and $B \in R^{n \times n}$ signifies the presence of chemical bonds within the respective drug molecule, alternatively is termed an intra-level graph. n refers to the atom numbers of the corresponding drug molecule. Here d_h is the dimensions of atom features and it indicates the dimensions of atom-level representations. θ is the parameters used in the molecular graph. Then the atom-level representations generated via the graph encoder $g_\theta(A, B)$ are expressed as $H = \{h_1, h_2..., h_n\}(H \in R^{n \times d_h})$.

DDI Relationship Network Graph. The DDI interaction relationship network can be denoted as $\mathcal{G} = (G, L)$ which is also named in inter-level representation. $G \in R^{N \times d_g}$ is denoted as the intra-level representation of drugs and $L \in R^{N \times N}$ denotes the link relationships between drugs. The whole inter-level descriptions about drugs is shown in $S = \{s_1, s_2..., s_N\} \quad (S \in R^{N \times d_g})$ computed from $g_\phi(G, L)$. N is the number of drugs in the DDI network, d_g is the dimension of drug representation and ϕ is the parameters used in the DDI relationships network graph.

Problem definition. Our aim is to anticipate drug interactions, a task that entails assessing the presence or absence of a linkage between two pharmaceutical compunds within the DDI network graph.

2.3 The DM-DDI Framework

Our poposed model, DM-DDI, integrates the intra-view drug representations through a GCN encoder (low-level graph) and the inter-view drug representations via another GCN encoder(high-level graph) to make an accuracy prediction of the interactions between random two drugs.

2.4 Feature Space Matching

Given a batch of n training samples of drug molecules $\{g_1, g_2, \cdots, g_n\}$ with rationales $\{d_1, d_2, \cdots, d_n\}$ computed by the bond-aware message passing networks and GCN, respectively. We obtain the intra-view features $\{g_1^\phi, g_2^\phi, \cdots, g_n^\phi\}$ and inter-view features $\{d_1^\varphi, d_2^\varphi, \cdots, d_n^\varphi\}$. A central moment discrepancy regularizer [40] is employed to match the distributions in the feature space:

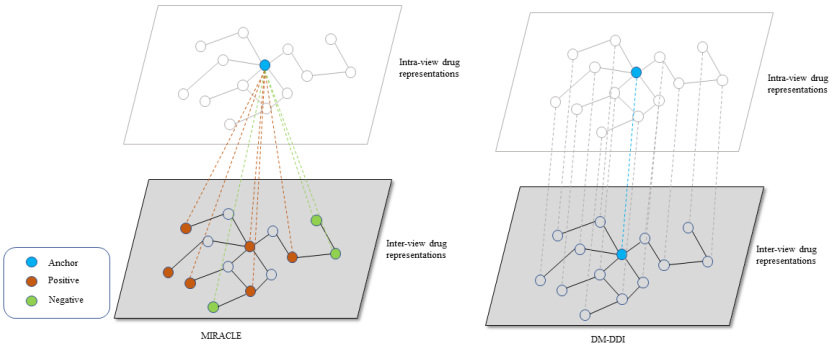

Fig. 3: The proposed graph learning framework compared with contrastive learning.

$$l_{fm} = \|E_{g^\phi} - E_{d^\varphi}\|_2 + \sum_{k=2}^{K} |C_k^{g^\phi} - C_k^{d^\varphi}\|_2 \tag{1}$$

$$E_{g^\phi} = \frac{1}{n}\sum_{i=1}^{n} g_i^\phi \tag{2}$$

$$E_{d^\varphi} = \frac{1}{n}\sum_{i=1}^{n} d_i^\varphi \tag{3}$$

$$C_k^{g^\phi} = \frac{1}{n}\sum_{i=1}^{n}(g_i^\phi - E_{g^\phi})^k \tag{4}$$

$$C_k^{d^\varphi} = \frac{1}{n}\sum_{i=1}^{n}(d_i^\varphi - E_{d^\varphi})^k \tag{5}$$

where the ϕ and φ are the set of parameters of BAMPN and GCN encoder. E_{g^ϕ} and E_{d^φ} are empirical expectation of the intra-view and inter-veiw features, and $C_k^{g^\phi}$ and $C_k^{d^\varphi}$ are the k-th order central moments of the feature coordinates. In practice, we compute the central moments up to the fifth order, i.e., $K = 5$. It is assumed that the features are distributed in the interval $[0, 1]$, the output of the sigmoid function; in other cases, constants might be added before each term.

2.5 Output Space Matching

We obtain two model distribution from inter-view interaction link embeddings and intra-view embeddings, namely p and r. We also need to make these consistent distribution between these two model distribution. The output space matching loss is written as the cross entropy between the intra-view distribution $r(Y)$ and the inter-view distribution $p(Y)$:

$$l_{om} = \sum_{y=1}^{|\mathcal{Y}|} -p_i(Y = y) \log r_i(Y = y) \tag{6}$$

Where Y is the prediction label of two model and y is the true label.

2.6 Overall Loss Function

Both predictors' primary goal is to minimize the supervised loss, which measures the distance between the predictions and the true labels. Another goal is to minimize a disagreement loss, which measures the distance between two predictors' predictions. The purpose of minimizing this disagreement loss is to enforce the model to pay more attention to the commonality between two different views and consistency between two predictors.

Formally, we formulate the supervised loss for the labeled interaction links and the disagreement loss for the unlabeled interaction links:

$$L_s = \sum_{l_i \in \mathcal{L}_l} (C(\mathbf{r_i}, \mathbf{y_i}) + \mathbf{C}(\mathbf{p_i}, \mathbf{y_i})) \tag{7}$$

$$L_d = \sum_{l_j \in \mathcal{L}_u} K(\mathbf{p_j} || \mathbf{r_j}) \tag{8}$$

where y_i is the true label of l_i, $\mathcal{L}_l$ and $\mathcal{L}_u$ denote the labeled and unlabeled links in $\mathcal{L}$ respectively, $C(\cdot, \cdot)$ is the cross-entropy loss function and $K(\cdot||\cdot)$ is the Kullback-Leibler divergence.

With feature sapce and outspace matching loss l_{fm} and l_{om}, supervised loss L_s and disagreement loss L_d, the objective function of our model is,

$$L = L_s + \alpha_1 l_{fm} + \alpha_2 l_{om} + \beta L_d \tag{9}$$

where α_1, α_2 and β are hyper-parameters for the trade-off for different loss components.

2.7 Drug-Drug Interaction Prediction

For each interaction link $l_{ij} \in \mathcal{L}$, We first compress two drug embedding vectors into an interaction link embedding vector:

$$l = d_i \odot d_j \tag{10}$$

where $\odot$ denotes the element-wise product, l is the interaction link embedding vector. Then, we apply a two-layer fully-connected neural network to make the final prediction:

$$p = \sigma(W_p\, ReLU(W_l l + b_l) + b_p) \tag{11}$$

where $p \in R^k$ and $[;]$ denotes the concatenation operation. If the aim is to predict the occurrence of DDI, k is 2.

Besides, we design another auxiliary interaction predictor using the inter-view drug embeddings. We cascade the last layer of the MPN with a fully connected layer and a sigmoid transformation function to construct this classifier. The prediction of the inter-view interaction predictor is denoted as $r \in R^k$. Optimizing the inter-view interaction prediction results can help the supervised information directly flow into previous network layers. The model learns the commonality between different views through a disagreement loss, which will be discussed below. So, finally, we get two prediction results corresponding to two different predictors. It should be noted that we only use p for the final DDI prediction.

2.8 Algorithmic Complexity Analysis

From the Figure 3, we can summarize up to that the MIRACLE algorithm used a contrastive learning to integrate the multi-view information into the prediction of DDI. In the process pf contrastive learning, positive samples and negative samples of one anchor node must be established and the loss compuatation will be $O(n^2)$, where n is the number of nodes. In our proposed model DM-DDI, element wise alignment scheme is used instead of the contrastive learning strategy. During the loss computation, we do not need to search positive and negative samples around every each anchor node, so that the computational complexity can be $o(n)$. The algorithm's low complexity in 50% improvement in computation time. With a slight increase in prediction performace, the ability to reduce computational speed is a highlight of this algotithm.

3 Experiments

In this section, we commence by presenting the datasets utilized shown in Table 1, the methods employed for comparison, and the evaluation metrics applied during the experiments. Subsequently, we conduct a comparative analysis of DM-DDI against other relevant methodologies. Finally, we perform an in-depth examination of DM-DDI's performance across a range of diverse experimental conditions.

3.1 Comparing Methods

In order to showcase the excellence of our proposed model, we have executed numerous baseline techniques to assess their performance. These baseline methods encompass both similarity-based and graph-based approaches. In the case of

the latter, to ensure a fair comparison of methods that utilize various perspectives of information, we have standardized them within the same framework as DM-DDI, as outlined in Table 2 and elaborated below.

• **Nearest Neighbor** [36]: Vilar and his team utilized drug interaction information and substructure-based similarity for DDI prediction, which we simply refer to as **NN**.

• **Label Propagation**: [37] employed the label propagation (LP) algorithm to construct three similarity-based predictive models. These models are based on substructure, side effects, and off-label side effects, respectively, and we refer to them as **LP-Sub**, **LP-SE** and **LP-OSE**, respectively.

• **Multi-Feature Ensemble**: [38] created a hybrid ensemble model by incorporating neighbor recommendation (NR), label propagation (LP), and matrix disturbs (MD) algorithms, which consider various aspects of drugs. We call this model **Ens**.

• **SSP-MLP**: [39] combined a pre-computed low-dimensional Structural Similarity Profile (SSP) with a Multi-layer Perceptron (MLP) for classification, and we label this model as **SSP-MLP**.

• **GCN**: [34] employed a graph convolutional network (GCN) for semi-supervised node classification tasks. We use **GCN** to encode drug molecular graphs and make predictions based on their representations as a baseline.

• **GIN**: [41] introduced a graph isomorphism network (GIN) for learning molecule representations in various single-property prediction tasks. We utilize **GIN** to encode drug molecular graphs and make predictions based on their representations as a baseline.

• **Attentive Graph Autoencoder**: [42] esigned an attentive mechanism to integrate multiple drug similarity views, which are then fed into a graph autoencoder to learn embedding vectors for each drug. This model is referred to as **AttGA**, and predictions are made based on the learned drug representations pairwise as a baseline.

• **GAT**: [43] utilized a graph attention network (GAT) to learn node embeddings through a well-designed attention mechanism on the graph. We apply **GAT** to obtain drug embeddings from the DDI network for predictions.

• **SEAL-CI**: [44] introduced a hierarchical graph representation learning framework for semi-supervised graph classification tasks, and we call this model **SEAL-CI**. We use it to learn drug representations for DDI predictions as a baseline.

• **NFP-GCN**: [45] developed the first graph convolution operator specifically designed for molecules. We refer to this model as **NFP-GCN** and incorporate it as a baseline by using its bond-aware message passing networks.

• **MIRACLE**: [46] developed a novel method which framework contains three sequential and independent phases, namely drug embeddings from the bond-aware message passing network, drug embeddings intrgration from multi-view graph representation using GCN encoder and contrastive learning-based strategy and interaction prediction. We refer to this model as **MIRACLE** (MultivIew gRAph Contrastive representation LEarning for drug-drug interaction prediction).

Table 2: Comparison of baseline methods.

Algorithm	Model Type	The Inter-view Model	The Intra-view Model	Feature Type
NN	similarity-based	N/A	N/A	similarity-based
LP	similarity-based	N/A	N/A	similarity-based fingerprint
Ens	similarity-based	N/A	N/A	similarity-based fingerprint
SSP-MLP	similarity-based	N/A	N/A	similarity-based fingerprint
GCN	inter-view	GCN	N/A	Molecular Graph
GIN	inter-view	GIN	N/A	Molecular Graph
AttGA	intra-view	N/A	AttGA	Interaction Relationship
GAT	intra-view	N/A	GAT	Interaction Relationship
SEAL-CI	multi-view	GCN	GCN	Molecular Graph & Interaction Relationship
NFP-GCN	multi-view	NFP	GCN	Molecular Graph & Interaction Relationship
MIRACLE	multi-view	BAMPN	GCN	Molecular Graph & Interaction Relationship
DM-DDI	multi-view	GCN	BAMPN	Molecular Graph & Interaction Relationship

3.2 Evaluation Metrics and Experimental Settings

We've split the interaction samples into a 4:1 ratio for training and testing, with a random 1/4 as a validation set. Our dataset contains only confirmed positive drug pairs, with randomly sampled negative pairs for training [47].

Parameter group learning rates decay exponentially from an initial rate of 0.0001. Hyperparameters include a 256-dimension hidden state, three bond-aware message passing neural networks, and a GCN encoder. Objective function coefficients α and β are set at 100 and 0.8, respectively, while dropout with $p = 0.3$ is applied to intermediate layers [48].

We implemented the model using Pytorch [49] and Pytorch-geometric 1.4.2 [50], utilizing the Adam [51] optimizer and Xavier [52] initialization. Model effectiveness is evaluated using *AUROC*, *AUPRC*, and *F1*, with results reported as mean and standard deviation over ten repetitions.

3.3 Experimental Results

We experimented with three distinct datasets to confirm the effectiveness of our proposed method in various scenarios. These experiments validate that our DM-DDI method outperforms the baselines in three different settings: a small-scale dataset with diverse drug features, a medium-scale dataset with limited labeled DDI links, and a large-scale dataset with missing drug features.

Comparison on the *ZhangDDI* dataset

Table 3 compares our DM-DDI model's performance against baseline approaches on the *ZhangDDI* dataset, which utilizes various drug features for DDI prediction. The best results are highlighted. DIM-DDI integrates multi-view information for drug representations, considering both inter-view drug molecular graphs and intra-view DDI relationships. According to the experiments, the proposed model outperforms the baseline approaches.

While algorithms using similarity-based fingerprints like **NN**, **LP**, and **SSP-MLP** performs poorly due to relying on only one type of features. **Ens** achieve better results by combining three models that use eight types of drug feature similarities and six topological features. This highlights the importance of integrating information from multiple sources like similarity-based fingerprints and topological features.

Some graph-based methods perform worse than the models mentioned above because they rely on the single view graph information. **GCN** and **GIN** encode drug molecular graphs using two different graph neural network frameworks for pairwise DDI prediction. **AttGA** and **GAT** directly learn drug representations from DDI interaction relationships. The former considers multiple connectivities of the DDI network and applies a GCN encoder to achive better performace than **GAT**, who only considers the existence of DDI network's links.

In multi-view graph settings, **SEAL-CI** and **NFP-GCN** and even **MIRACLE** outperform other baselines, showing that the integration of multi-view graph can significantly enhance model performance. However, their performance is still not as strong as the proposed DDI method in addition to the MIRACLE. MIRACLE and DDI futher considers the importance of the message passing within chemical bonds inside drug molecular graphs and maintains a balance between multi-view graph information to learn more comprehensive drug representations. In contrast, **SEAL-CI** and **NFP-GCN** explicitly model multi-view graphs but overlook the information equilibrium beteween different views. Additionally, **DM-DDI** uses a self-attentive mechanism to generate inter-view drug representations by selecting the most significant atoms for meaningful functional groups in DDI reactions while ignoring noisy, meaningless substructures. However, compared with the **MIRACLE**, our method **DM-DDI** achive better performace on the metrics AUPRC and F1 Score with additional feature space matching and output space matching in a more efficient way.

Comparison on the *ChCh-Miner* dataset

In this experimental section, we aim to assess the performance of our pro-

Table 3: Comparative evaluation results on *ZhangDDI*

Algorithm	Performance		
	AUROC	AUPRC	F1
NN	67.81 ± 0.25	52.61 ± 0.27	49.84 ± 0.43
LP-Sub	93.39 ± 0.13	89.15 ± 0.13	79.61 ± 0.16
LP-SE	93.48 ± 0.25	89.61 ± 0.19	79.83 ± 0.61
LP-OSE	93.50 ± 0.24	90.31 ± 0.82	80.41 ± 0.51
Ens	95.20 ± 0.14	92.51 ± 0.15	85.41 ± 0.16
SSP-MLP	92.51 ± 0.15	88.51 ± 0.66	80.69 ± 0.81
GCN	91.91 ± 0.62	88.73 ± 0.84	81.61 ± 0.39
GIN	81.45 ± 0.26	77.16 ± 0.16	64.15 ± 0.16
AttGA	92.84 ± 0.61	90.21 ± 0.19	70.96 ± 0.39
GAT	91.49 ± 0.29	90.69 ± 0.10	80.93 ± 0.25
SEAL-CI	92.93 ± 0.19	92.82 ± 0.17	84.74 ± 0.17
NFP-GCN	93.22 ± 0.09	93.07 ± 0.46	85.29 ± 0.38
MIRACLE	$\mathbf{98.95 \pm 0.15}$	$\mathbf{98.17 \pm 0.06}$	$\mathbf{93.20 \pm 0.27}$
DM-DDI	**98.42**	**96.31**	**92.72**

posed **DM-DDI** method on datasets with limited labeled DDI links. We only compare **DM-DDI** with graph-based baselines because this dataset lacks similarity-based fingerprints for drug pairs. Due to this limitation, **AttGA** is not applicable to this dataset. The results are presented in Table 4.

Clearly, methods that consider multi-view information, such as **SEAL-CI** ,**NFP-GCN**, **MIRACLE** and **DM-DDI** outperform baselines that rely on single-view information. However, **DM-DDI** outhsines them all, demonstrating its superiority on datasets with scarce labeled data.

The graph contrastive learning component in **MIRACLE** effectively integrates and balances information from different views, enabling **MIRACLE** learns drug representations better with limited labeled DDI links. However, the compuation cost of **MIRACLE** is expensive compared with our proposed model **DM-DDI**. The proposed method **DM-DDI** also considers integration and balance information from intra-view and inter-view using the feature and output space matching, considering the DDI network's structural information when making the final predicting decisions, which helps the model extracts the most useful information from all dimensional features for DDI prediction. We further verify our points by a small ablation study in which we adjust the training ratio of the dataset in subsection 3.4.

Comparison on the *DeepDDI* dataset

To assess the scalability of our proposed method, we conducted experiments on a large-scale dataset, *DeepDDI*, which contains abundant labeled data and DDI information. In Table 7, we compare the performance of our approach to

Table 4: Comparative evaluation results on *ChCh-Miner*

Algorithm	Performance		
	AUROC	AUPRC	F1
GCN	82.84 ± 0.61	84.27 ± 0.66	70.54 ± 0.87
GIN	70.32 ± 0.87	72.41 ± 0.63	65.54 ± 0.97
GAT	85.84 ± 0.23	88.14 ± 0.25	76.51 ± 0.38
SEAL-CI	90.93 ± 0.19	89.38 ± 0.39	84.74 ± 0.48
NFP-GCN	92.12 ± 0.09	93.07 ± 0.69	85.41 ± 0.18
MIRACLE	$\mathbf{96.15 \pm 0.29}$	$\mathbf{95.57 \pm 0.19}$	$\mathbf{92.26 \pm 0.09}$
DM-DDI	**96.59**	**97.94**	**97.82**

the baseline methods. Many baseline approaches rely on similarity-based fingerprints, which often require a siginificant number of non-structural similarity features, which may be lacking in this large scale dataset. Therefore, we only consider models that are applicable to this dataset, including **NN** and **SSP-MLP**. Among the graph-based methods, we exclude **AttGA** due to the absence of many necessary drug features. Additionally, We do not present experimental results for **GIN** and **NFP-GCN** due to their inferior performance and space limitation.

MLP-SSP outperforms **NN**, primarily because the former framework is based on deep neural networks. **GCN** achieves better results than **GAT**, further underscoring the importance of inter-view information in DDI predictions. Among the baselines, **SEAL-CI** follows follows the proposed method **DM-DDI** model closely, emphasizing the strength of the multi-view graph framework. **MIRACLE** is the method using contrastive learning in the integration information from different view. **DM-DDI** using another efficient way in the distribution matching and significantly outperformed other baseline methods in terms of all the three metrics.

Table 5: Comparative evaluation results on *DeepDDI*

Algorithm	Performance		
	AUROC	AUPRC	F1
NN	81.81 ± 0.37	80.82 ± 0.20	71.37 ± 0.18
SSP-MLP	92.28 ± 0.18	90.27 ± 0.28	79.71 ± 0.16
GCN	85.53 ± 0.17	83.27 ± 0.31	72.18 ± 0.22
GAT	84.84 ± 0.23	81.14 ± 0.25	73.51 ± 0.38
SEAL-CI	92.83 ± 0.19	90.44 ± 0.39	80.70 ± 0.48
MIRACLE	$\mathbf{95.51 \pm 0.27}$	$\mathbf{92.34 \pm 0.17}$	$\mathbf{83.60 \pm 0.33}$
DM-DDI	**93.17**	**96.48**	**95.82**

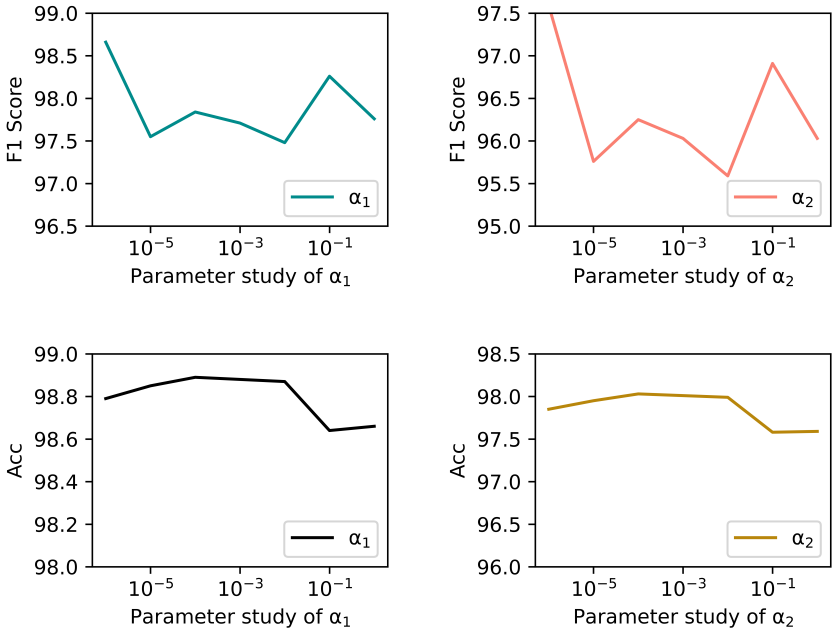

Fig. 4: Parameter sensitivity study on *CHMiner*

3.4 Ablation Study

Different matching loss comparision

To achieve kinds of different matching loss comparison, we conducted experiments on DM-DDI based on different matching loss using *CHMiner* datasets. The results are shown in Table 5. We can see our DMR matching loss outperformed than the coral, MMD, GAN, JSD and DV matching loss.

Table 6: Comparision of Different Matching Loss on *CHMiner*

Method	Performance	
	F1	ACC
DMR-coral	97.838	96.245
DMR-MMD	97.84	96.25
DMR-GAN	97.75	96.08
DMR-JSD	97.83	96.23
DMR-DV	98.38	97.11
DMR-CMD	98.66	97.59

We conduct ablation experiments on the *ChCh-Miner* dataset to validate the feature space matching and output space matching component's effectiveness and in our **DM-DDI** model. The experimental results are reported in Table 4. To better understand the differences between DM-DDI with and without the feature and output space matching component,we can observe that DM-DDI effectively

learns drug embeddings with conclusion two space matching components. DM-DDI with the feature and output space matching component can also achieve high performance.

Table 7: Ablation experimental results on fm & om. "-fm" means removing the feaure space matching loss, and "-fm & om" means removing both reature and output space matching losses.

Method	Performance	
	F1	ACC
DM-DDI	98.66	97.59
-fm	98.322	96.99
-fm and -om	98.38	97.11

Parameter Sensitivity In our model in equation 9, there are two major parameters α_1 and α_2. In this subsection, we evaluate the impacts of them, together with the dimensionality of drug embeddings d_g on the *CHMiner* dataset. Figure **??** to **??** show the results by changing one parameter while fixing another one.

First, we vary α_1 by {*1e-1,1e-2,1e-3,1e-4,1e-5,1e-6*}, and fix $\alpha_2 = 0.8$. Here, for parameter study purpose, α_2 is set to its optimal value on this dataset instead of the default value 1. From the figure, our method is quite stable in a wide range of α_1 and achieves the best performance when $\alpha = 1e-6$ in terms of *F1 Score*. Next, we vary α_2 by {*8e-1, 8e-2, 8e-3, 8e-4, 8e-5, 8e-6*} with $\alpha_1 = 1e-6$ and $d_g = 256$. As can be seen, the near optimal performance at $\alpha_2 = 8e-4$ justifies our parameter setting. Overall, α_1 and α_2 are stable w.r.t these parameters. Moreover, the non-zero choices of α_1 and α_2 demonstrate the importance of the loss terms in our model.

We also take a parameter (α_1 and α_2) test on the metric of Acc. As can be seen, the near optimal performance at α_1=1e-6 and $\alpha_2 = 8e-4$ justifies our parameter setting. Overall, α_1 and α_2 are stable w.r.t these parameters. Moreover, the non-zero choices of α_1 and α_2 demonstrate the importance of the loss terms in our model.

ylabel=Objective function value, xmin=-10, xmax=610, ymin=-450, ymax=2100, xtick=0,150,300,450,600, ytick=-400,200,800,1400,2000, legend pos=north east, ymajorgrids=true, grid style=dashed, no markers, every axis plot/.append style=thick

4 conclusion

To effectively harness the rich multi-view graph information within the DDI network, we introduce **DM-DDI** for the task of DDI prediction in this paper. **DM-DDI** adopts a multi-view graph perspective for learning drug embeddings, achieved trough an end-to-end framework comprising a bond-aware message passing network and a GCN encoder. Furthermore, we introduce a novel

distribution matching approach to balance information from different views. Additionally, the DM-DDI also designs two predictors based on both views to maximize the utilization of available information. Through extensive experimentation with various datasets, we have substantiated that DM-DDI is not only highly effective but also efficient in its DDI prediction capabilities include computation cost.

References

1. Wani, M.Y., Ahmad, A.: Combination Therapy Against Multidrug Resistance. Academic Press (2020)
2. Palleria, C., Di Paolo, A., Giofrè, C., Caglioti, C., Leuzzi, G., Siniscalchi, A., De Sarro, G., Gallelli, L.: Pharmacokinetic Drug-Drug Interaction and Their Implication in Clinical Management. Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences **18**(7), 601 (2013)
3. Wang, Y., Xu, Y., Yang, Z., Liu, X., Dai, Q.: Using Recursive Feature Selection with Random Forest to Improve Protein Structural Class Prediction for Low-Similarity Sequences. Computational and Mathematical Methods in Medicine **2021** (2021)
4. Dai, Q., Bao, C., Hai, Y., Ma, S., Zhou, T., Wang, C., Wang, Y., Huo, W., Liu, X., Yao, Y.: MTGIpick Allows Robust Identification of Genomic Islands from a Single Genome. Briefings in Bioinformatics **19**(3), 361–373 (2018)
5. Yang, S., Wang, Y., Chen, Y., Dai, Q.: MASQC: Next Generation Sequencing Assists Third Generation Sequencing for Quality Control in N6-methyladenine DNA Identification. Frontiers in Genetics **11**, 269 (2020)
6. Kong, R., Xu, X., Liu, X., He, P., Zhang, M.Q., Dai, Q.: 2SigFinder: The Combined Use of Small-Scale and Large-Scale Statistical Testing for Genomic Island Detection from a Single Genome. BMC Bioinformatics **21**(1), 1–15 (2020)
7. Yang, Z., Yi, W., Tao, J., Liu, X., Zhang, M.Q., Chen, G., Dai, Q.: HPVMD-C: A Disease-Based Mutation Database of Human Papillomavirus in China. Database **2022**, baac018 (2022)
8. Dara, S., Dhamercherla, S., Jadav, S.S., Babu, C.H.M., Ahsan, M.J.: Machine Learning in Drug Discovery: A Review. Artificial Intelligence Review **55**(3), 1947–1999 (2022)
9. Askr, H., Elgeldawi, E., Aboul Ella, H., Elshaier, Y.A.M.M., Gomaa, M.M., Hassanien, A.E.: Deep Learning in Drug Discovery: An Integrative Review and Future Challenges. Artificial Intelligence Review **56**(7), 5975–6037 (2023)
10. Vilar, S., Uriarte, E., Santana, L., Lorberbaum, T., Hripcsak, G., Friedman, C., Tatonetti, N.P.: Similarity-Based Modeling in Large-Scale Prediction of Drug-Drug Interactions. Nature Protocols **9**(9), 2147–2163 (2014)
11. Cheng, F., Zhao, Z.: Machine Learning-Based Prediction of Drug-Drug Interactions by Integrating Drug Phenotypic, Therapeutic, Chemical, and Genomic Properties. Journal of the American Medical Informatics Association **21**(e2), e278–e286 (2014)
12. Zhang, P., Wang, F., Hu, J., Sorrentino, R.: Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects. Scientific Reports **5**(1), 12339 (2015)
13. Kastrin, A., Ferk, P., Leskošek, B.: Predicting Potential Drug-Drug Interactions on Topological and Semantic Similarity Features Using Statistical Learning. PLOS ONE **13**(5), e0196865 (2018)

14. Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P.: Convolutional Networks on Graphs for Learning Molecular Fingerprints. Advances in Neural Information Processing Systems **28** (2015)
15. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural Message Passing for Quantum Chemistry. In: International Conference on Machine Learning, pp. 1263–1272. PMLR (2017)
16. Kipf, T.N., Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks. arXiv preprint arXiv:1609.02907 (2016)
17. Asada, M., Miwa, M., Sasaki, Y.: Enhancing Drug-Drug Interaction Extraction from Texts by Molecular Structure Information. arXiv preprint arXiv:1805.05593 (2018)
18. Huang, Y., Chen, Y., Du, Y., Yang, Z.: Distribution matching for rationalization. Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 14, pp. 13090–13097 (2021).
19. Chen, X., Liu, X., Wu, J.: GCN-BMP: Investigating Graph Representation Learning for DDI Prediction Task. Methods **179**, 47–54 (2020). Elsevier
20. Mehrjardi, F.Z., Latif, A.M., Zarchi, M.S., Sheikhpour, R.: A Survey on Deep Learning-Based Image Forgery Detection. Pattern Recognition, 109778 (2023). Elsevier
21. Li, J., Rong, Y., Cheng, H., Meng, H., Huang, W., Huang, J.: Semi-supervised graph classification: A hierarchical graph perspective. In: F., E., S., E. (eds.) The World Wide Web Conference, pp. 972–982 (2019).
22. Li, J., Huang, Y., Chang, H., Rong, Y.: Semi-Supervised Hierarchical Graph Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022). IEEE.
23. Bai, Y., Gu, K., Sun, Y., Wang, W.: Bi-level graph neural networks for drug-drug interaction prediction. arXiv preprint arXiv:2006.14002 (2020).
24. Wang, Y., Min, Y., Chen, X., Wu, J.: Multi-view graph contrastive representation learning for drug-drug interaction prediction. In: F., E., S., E. (eds.) Proceedings of the Web Conference 2021, pp. 2921–2933 (2021).
25. He, C., Liu, Y., Li, H., Zhang, H., Mao, Y., Qin, X., Liu, L., Zhang, X.: Multi-type feature fusion based on graph neural network for drug-drug interaction prediction. BMC Bioinformatics, vol. 23, no. 1, pp. 224 (2022). Springer.
26. Huang, Y., Chen, Y., Du, Y., Yang, Z.: Distribution matching for rationalization. In: F., E., S., E. (eds.) Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 14, pp. 13090–13097 (2021).
27. Kullback, S., Leibler, R.A.: On information and sufficiency. The Annals of Mathematical Statistics, vol. 22, no. 1, pp. 79–86 (1951). JSTOR.
28. Wang, Y., Min, Y., Chen, X., Wu, J.: Multi-view Graph Contrastive Representation Learning for Drug-Drug Interaction Prediction. In: F., E., S., E. (eds.) Proceedings of the Web Conference 2021, ACM (2021). https://doi.org/10.1145/3442381.3449786.
29. Huang, Y., Chen, Y., Du, Y., Yang, Z.: Distribution matching for rationalization. In: F., E., S., E. (eds.) Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 14, pp. 13090–13097 (2021).
30. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural Message Passing for Quantum Chemistry. In: Precup, D., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017, vol. 70, pp. 1263–1272. PMLR (2017). URL: http://proceedings.mlr.press/v70/gilmer17a.html.

31. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: F., E., S., E. (eds.) Advances in Neural Information Processing Systems, pp. 5998–6008 (2017).
32. Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks. In: ICML 2015 Deep Learning workshop (2015). URL: https://arxiv.org/abs/1505.00387.
33. Li, Y., Tarlow, D., Brockschmidt, M., Zemel, R.S.: Gated Graph Sequence Neural Networks. In: Bengio, Y., LeCun, Y. (eds.) 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings (2016). URL: http://arxiv.org/abs/1511.05493.
34. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016).
35. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, pp. 40–51 (2006). IEEE.
36. Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., Friedman, C.: Drug—drug interaction through molecular structure similarity analysis. Journal of the American Medical Informatics Association, vol. 19, no. 6, pp. 1066–1074 (2012). BMJ Group.
37. Zhang, P., Wang, F., Hu, J., Sorrentino, R.: Label propagation prediction of drug-drug interactions based on clinical side effects. Scientific Reports, vol. 5, no. 1, pp. 1–10 (2015). Nature Publishing Group.
38. Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G., Li, X.: Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC Bioinformatics, vol. 18, no. 1, pp. 18 (2017). Springer.
39. Ryu, J.Y., Kim, H.U., Lee, S.Y.: Deep learning improves prediction of drug–drug and drug–food interactions. Proceedings of the National Academy of Sciences, vol. 115, no. 18, pp. E4304–E4311 (2018). National Acad Sciences.
40. Zellinger, W., Grubinger, T., Lughofer, E., Natschläger, T., Saminger-Platz, S.: Central moment discrepancy (CMD) for domain-invariant representation learning. arXiv preprint arXiv:1702.08811 (2017).
41. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How Powerful are Graph Neural Networks? In: Bengio, Y., LeCun, Y. (eds.) 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net (2019). URL: https://openreview.net/forum?id=ryGs6iA5Km.
42. Ma, T., Xiao, C., Zhou, J., Wang, F.: Drug Similarity Integration Through Attentive Multi-view Graph Auto-Encoders. In: Lang, J. (ed.) Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden, pp. 3477–3483. ijcai.org (2018). https://doi.org/10.24963/ijcai.2018/483.
43. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph Attention Networks. In: Bengio, Y., LeCun, Y. (eds.) 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net (2018). URL: https://openreview.net/forum?id=rJXMpikCZ.
44. Li, J., Rong, Y., Cheng, H., Meng, H., Huang, W., Huang, J.: Semi-supervised graph classification: A hierarchical graph perspective. In: F., E., S., E. (eds.) The World Wide Web Conference, pp. 972–982 (2019).
45. Duvenaud, D.K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P.: Convolutional networks on graphs for learning molecular fingerprints. In: F., E., S., E. (eds.) Advances in Neural Information Processing Systems, vol. 28 (2015).

46. Wang, Y., Min, Y., Chen, X., Wu, J.: Multi-view Graph Contrastive Representation Learning for Drug-Drug Interaction Prediction. In: F., E., S., E. (eds.) Proceedings of the Web Conference 2021, pp. 2921–2933 (2021).
47. Chen, X., Liu, X., Wu, J.: Drug-drug Interaction Prediction with Graph Representation Learning. In: F., E., S., E. (eds.) 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 354–361. IEEE (2019).
48. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958 (2014). JMLR.org.
49. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: F., E., S., E. (eds.) Advances in Neural Information Processing Systems, pp. 8026–8037 (2019).
50. Fey, M., Lenssen, J.E.: Fast Graph Representation Learning with PyTorch Geometric. In: F., E., S., E. (eds.) ICLR Workshop on Representation Learning on Graphs and Manifolds (2019).
51. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. URL: http://arxiv.org/abs/1412.6980 (2015).
52. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: F., E., S., E. (eds.) Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010).