

# 基于图卷积网络的药物与药物相互作用预测的研究进展

高鸿雁

辽宁科技大学计算机与软件工程学院, 辽宁 鞍山

收稿日期: 2022年3月18日; 录用日期: 2022年4月15日; 发布日期: 2022年4月21日

## 摘要

药物之间的相互作用(DDIs)可能会导致严重的医疗伤害, 确定DDIs并探讨其潜在机制对复合药安全至关重要。然而, 在现实中由于需要对大量药物组合进行实验研究, 分析检测DDIs仍然费时且昂贵。大多数药物相关知识是临床评估和上市后监测的结果, 导致信息的数量有限。正确预测药物-药物相互作用不仅可以减少这些病例, 还可以降低药物开发成本。因此, 开发了许多计算方法来预测DDIs。其中基于图卷积网络的模型被提出用来预测DDIs。并在药物相互作用预测方面取得了最先进的结果。在此我们回顾了近年来基于图卷积网络的药物相互作用预测的研究现状, 调查了近期研究人员在实验中利用到的最新方法和数据库。实验结果表明, 基于图卷积网络的预测模型优于其他基于深度学习的算法。证明了基于图卷积网络来预测药物之间相互作用确实是有必要的和有价值的。

## 关键词

药物, 相互作用, 图卷积网络, DDIs

# Research Progress on Predicting Interaction between Drugs Based on Graph Convolution Network

Hongyan Gao

School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan Liaoning

Received: Mar. 18<sup>th</sup>, 2022; accepted: Apr. 15<sup>th</sup>, 2022; published: Apr. 21<sup>st</sup>, 2022

## Abstract

Drug-drug interactions (DDIs) may lead to serious medical injuries. It is important to identify DDIs

and investigate the underlying mechanisms of DDIs for the safety of compound drugs. However, in practice, analytical detection of DDIs is still time-consuming and expensive due to the need for experimental studies on a large number of drug combinations. Most drug-related knowledge is the result of clinical evaluation and post-marketing surveillance, resulting in a limited amount of information. Correctly predicting drug-drug interactions could not only reduce these cases but also reduce the cost of drug development. Therefore, a number of computational methods have been developed to predict DDIs. The model based on graph convolution network is proposed to predict DDIs and has achieved the most advanced results in drug interaction prediction. Here we review the research status of drug interaction prediction based on graph convolutional networks in recent years and investigate the latest methods and databases used by researchers in recent experiments. Experimental results show that the prediction model based on graph convolution network outperforms other algorithms based on deep learning. It is proven that it is necessary and valuable to predict drug-drug interactions based on graph convolutional networks.

## Keywords

Drugs, Interaction, Graph Convolutional Network, DDIs

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 介绍

近年来,多药联用正成为治疗复杂疾病(如糖尿病和癌症)的一种有前途的策略。当两种或两种以上药物同时服用时,可能会引发意想不到的副作用、不良反应、甚至严重的毒性反应。多种药物在治疗过程中引发的药理作用称为药物-药物相互作用(drug-drug interactions, DDIs)。因此,重要的是确定 DDIs 并揭示其对多药安全的潜在机制。然而,在大量药物对的检测中发现 DDIs 既费时又昂贵,所以很多计算方法被提出用于预测 DDIs。在过去的十年中,这种通过实验确定 DDIs 促进了计算方法的应用,特别是通过深度学习的计算方法以发现潜在的 DDIs。各种基于深度学习的计算方法伴随着效率高、成本低的优点已经被证明是一种很有前途的方法,可以提供 DDIs 的初步筛选。在本篇文章中,我们着重介绍近年来基于图卷积网络的药物相互作用预测及其最新研究进展,这是其发挥生物医学功能的一个重要作用机制。

## 2. 通用数据集

目前,药物相互作用的可靠的数据分布在多个来源上。已知的主流药物相互作用数据库是 DrugBank 数据库,此外还有 PharmGKB、KEGG Drug、OFFSIDES 和 BioSNAP-DDIs。DrugBank 数据库是一个权威标准 DDIs 数据库,涵盖 294,980 个 DDIs (阳性),跨 2,286 个批准药物。它结合了详细的药物相关信息,包括化学、药理、药物数据和全面的药物靶标信息。PharmGKB 数据库包含 12,664 种药物条目,包括 2,588 种已批准的小分子药物、1,287 种已批准的生物技术药物、130 种营养药物和超过 6,305 种实验药物。PharmGKB 数据库是一个存储基因组、分子和细胞表型数据的仓库。它还包含临床信息和遗传变异对药物反应的影响。PharmGKB 包含基因、疾病、药物和通路相关数据,以及影响药物代谢的 470 个遗传变异的详细信息。KEGGDrug 数据库包含代谢途径,链接到代谢产物和蛋白质/酶信息。截至 2019 年 5 月,KEGGDrug 数据库拥有药品相关信息 10,979 条,DDIs 关系 501,689 条。OFFSIDES 数据库包含了基于 PharmGKB 的不良事件报告中挖掘出来的药物效应,该数据库报告了 438,802 种药物副作用。

最后, BioSNAP-DDIs 数据库包含从药物标签和生物医学文献中提取的 1,514 种药物之间的一共 48,514 次相互作用。

### 3. 图卷积网络模型介绍

近年来, 卷积网络快速发展[1] [2], 并因其建模的高效性引起大家重视。卷积网络给图像处理[3], 图像识别[4], 机器翻译[5], 语音识别[6]带来了很大的突破。传统卷积神经网络局限于对欧氏空间数据进行建模, 而图卷积操作利用图表示对非欧氏空间数据进行处理, 使卷积操作更加适用于药物数据结构。图卷积操作主要包括基于空域以及频域的图卷积类。空域图卷积将图卷积操作定义为图中相邻节点之间的特征信息的聚合, 基于频域的图卷积则利用图信号处理(graph signal processing), 引入滤波器实现频域图卷积的推导。无论是空域图卷积还是频域图卷积, 本质上都是对图拓扑结构的建模利用, 即空间特征的提取[7]。

首先给出常见通用的符号定义, 如表 1 所示。

**Table 1.** Common symbol definitions  
**表 1.** 常用的符号定义

符号	表示含义
$G$	图
$V$	节点集
$v$	节点
$n$	节点数
$E$	边集
$e$	边
$I_n \in R^{n \times n}$	单位矩阵
$A \in R^{n \times n}$	邻接矩阵
$H \in R^{n \times e}$	关联矩阵
$W \in R^{n \times n}$	权重矩阵
$D_e \in R^{n \times n}$	边度矩阵
$D_v \in R^{n \times n}$	节点度矩阵

一般地,  $c = (v; E; w)$  用于表示常用的无向带权图。 $A$  表示图的邻接矩阵(adjacency matrix), 定义节点之间的相互连通关系。 $H$  为关联矩阵(incidence matrix), 用于定义图中边与节点之间的连接, 当  $v_j \in e_k$ ,  $h_{jk} = 1$ , 其他情况下,  $h_{jk} = 0$ , 对于图中的节点, 其度矩阵  $D_v$  中元素定义为连接到该节点所有边的权重之和[7]。即

$$d(v) = \sum_{e \in E} W(e_i) h(v, e_i), \quad i \in I$$

相似地,  $D_e$  中元素定义为连接到该边的节点数量之和, 即

$$d(v) = \sum_{v \in V} h(v, e_i), \quad i \in I$$

GCN 的基本思路: 对于每个节点, 我们从它的所有邻居节点处获取其特征信息, 这当然也包括它本身的特征。将从邻居节点获得的特征信息和本身的特征信息融合, 输入到 average() 函数中, 得到一个平

均值。同样的，我们再对其他节点进行类似的操作。最后，我们将这些计算得到的所有平均值输入到神经网络中。

#### 4. 预测模型

在 2022 年，Wang 等人提出一种多核图卷积网络(GCNMK)来预测潜在的 DDI 的 [8]。该模型将 DDI 图分为两大类，其中一类是“增加”相关的图 GI，例如，药物 A 可能会增加药物 B 的活性。另外一类与“减少”有关的图 GD，例如，当药物 A 与药物 B 结合时药物 A 的代谢会降低。然后将两类 DDI 图 GD 和图 GI 分别与药物特征矩阵整合，分别输入两个 GCN 块。GCN 层有两层，第  $L_1$  层有两个并列的 GCN 块，第  $L_2$  层有一个 GCN 块。第  $L_2$  层的 GCN 块被用于去整合第一层的两个 GCN 块的输出，第  $L_2$  层的 GCN 块输出的是最终 DDI 特征矩阵，是药物的低维表示向量。在每个块中采用的是输入的线性组合的加法程序。最后将 DDI 特征矩阵输入三个完全连通层来预测潜在的 DDI 的。在隐藏层的激活函数用的是 RELU，输出层用的是 Sigmoid，损失函数用的是 cross-entropy 和 L2-regularization。L2-regularization 用来解决过拟合问题，损失函数越小，预测的准确性越高。实验调整了参数去实现最优的性能，最后计算最优的参数为学习率  $lr = 0.002$ ，L2-regularization 系数  $\lambda = 0.0003$ 。实验最后通过五倍交叉验证，证明了该模型确实可以有效进行药物和药物相互作用关系预测。具体流程如图 1 所示。

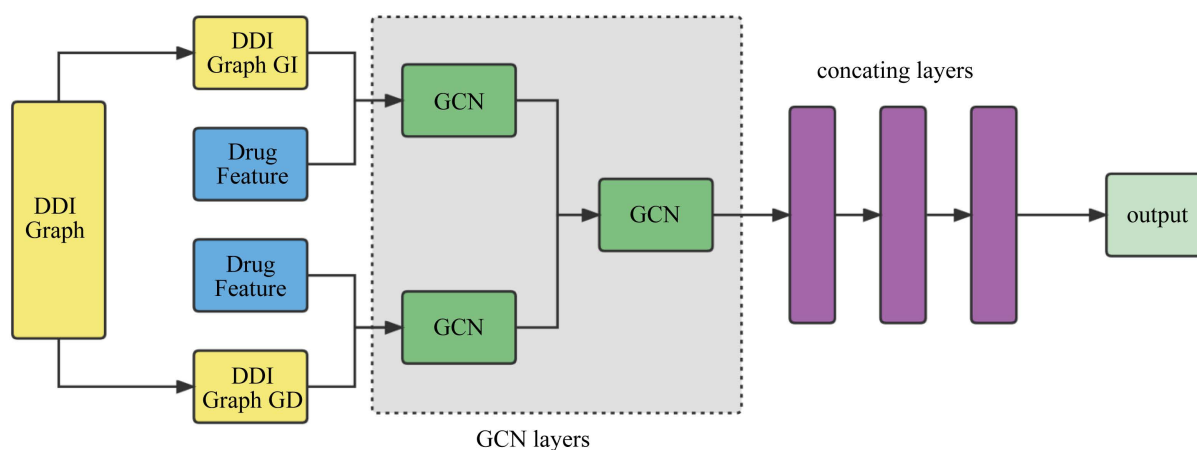


Figure 1. Flow chart of GCNMK model

图 1. GCNMK 模型流程图

在 2022 年，Feng 等人提出一种用于药物-药物相互作用多标签预测的深度图卷积网络框架(称为 deepMDDIs)来预测 DDI 及其类型[9]。首先将多关系网络 DDI 分解成  $m$  个子网络，用  $m$  个邻接矩阵表示子网络去作为编码器的输入。deepMDDIs 利用相似正则化约束，用深度图卷积网络设计了一种编码器来捕获 DDI 网络的拓扑特征。编码器通过捕捉药物的复杂拓扑性质，构造一个  $p$  层关系 GCN (R-GCN)，将多关系 DDI 网络中的药物编码为嵌入向量，并且从第 2 个隐藏层添加一个残留策略到最后一个隐藏层。同时，采用药物相似度矩阵约束嵌入空间的相似药物，使其尽可能接近。使一个节点在网络中聚合其一阶邻居节点以及自身的特征，以重新更新自己特征。然后，将所有更新的特征累加起来，通过一个 ReLU 激活函数，生成最终的嵌入。整个 DDI 网络由一个  $p$  层的 R-GCN 进行传播，以捕获其  $p$  阶邻居的信息。之后采用类张量解码器对单次交互和多次交互进行统一建模，以确定未标记的类型特异性药物对是否产生一种或多种药理作用。译码器是一种类张量分解矩阵运算，将嵌入的特征矩阵、特定类型的特征重要性矩阵和平均的特征关联矩阵相结合，重构出多元关系的 DDI 网络。

通过实验结果表明, deepMDDIs 可以有效区分 DDIs 网络中一个未标记的类型特异性药物对是否导致一种或多种药理作用, 也可以预测 DDIs 网络中已知药物与网络外新药之间的潜在 DDIs。通过与目前最先进的基于深度学习的方法在 DDIs 多分类任务中的比较, 在单倍 DDIs 和多倍 DDIs 预测中均取得了令人鼓舞的效果。证明了该算法的优越性。具体流程如图 2 所示。

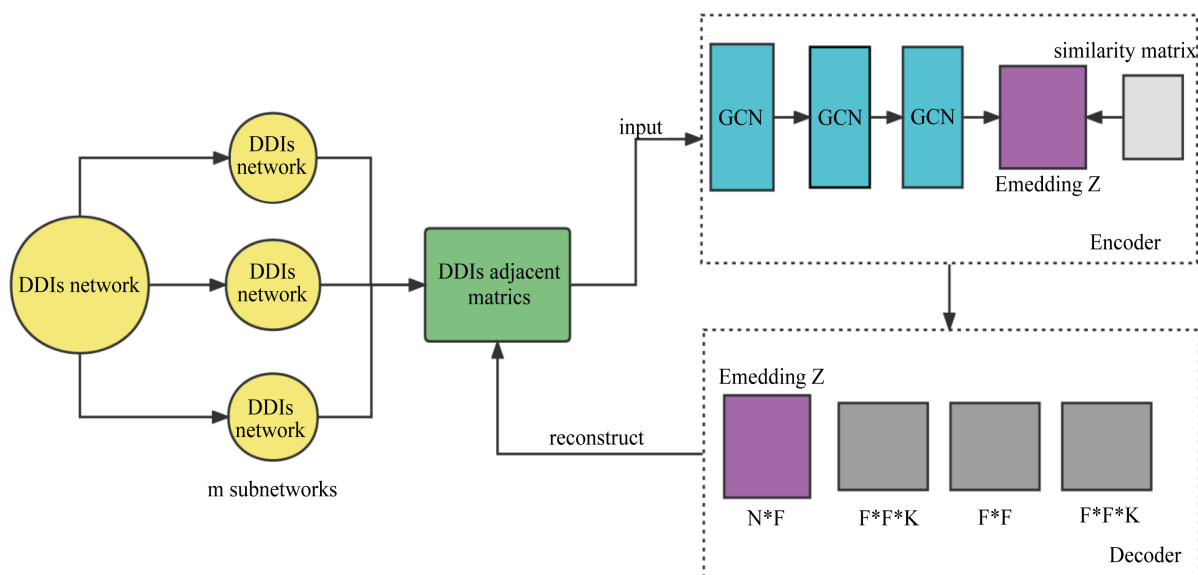


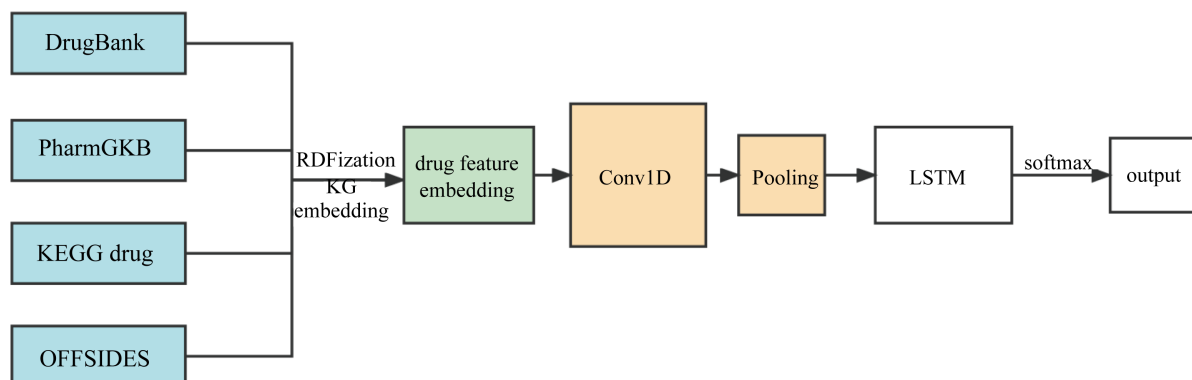
Figure 2. Flow chart of deepMDDIs model  
图 2. deepMDDIs 模型流程图

在 2019 年, Karim 等人提出了一种基于知识图嵌入和 Conv-LSTM 网络的药物 - 药物相互作用预测[10]。使用来自 DrugBank、KEGG Drug、PharmGKB 和 OFFSIDES 的数据创建了一个包含 2,898,937 个药物 - 药物交互对的数据集, 由于该背景数据的格式不能作为典型分类器的直接输入, 因此应用了几种节点嵌入技术, 使用 RDF2Vec、SimpleIE、TransE、KGloVe、CrossE 和 PBG 进行 KG 嵌入, 为 KG 中的每个节点创建一个密集的向量表示, 通过执行 KG 嵌入程序将信息从图编码成密集向量。KG 嵌入包括三个步骤: 表示实体和关系、定义评分函数以及学习实体和关系表示。之后将 KG 嵌入后的结果输入由 CNN 和 LSTM 层构造的 Conv-LSTM 网络中, CNN 使用卷积滤波器来捕捉药物特征中的局部关系值, 而 LSTM 网络可以从 CNN 提取的特征中携带所有关系。

网络中第一层是嵌入层, 它将药物样本作为一个序列映射到一个实向量域。然后将大小为  $100 \times 300$  的嵌入矩阵送入一维卷积层, 该层有 100 个滤波器, 核大小为 4。然后将每个卷积层的输出传递给 Dropout 层进行正则化学习以避免过拟合。直观上可以被认为是迫使分类器不依赖于任何微不足道的单个药物特征。然后卷积层将输入特征空间卷积为  $100 \times 100$  表示, 由沿着嵌入维的池大小为 4 的 1 维最大池化层 (MPL) 进行下一步采样, 产生形状为  $25 \times 100$  的输出。其中, 25 个维度中的每一个都可以被认为是一个提取的特征。MPL 通过在每个时间步长维度上取最大值来扁平化输出空间, 从而产生一个  $1 \times 100$  的载体。LSTM 层将扁平的特征向量的维度视为时间步长, 并且每个时间步长输出 100 个隐藏的单元。第二个 LSTM 层发出一个输出 H, H 被重塑成一个特征序列, 并被送入完全连接的层, 以预测下一步的 DDIs。然后使用一个全局 MPL, 在经过另一个 Dropout 后, 最具影响力的特征被送入一个全连接层, 最后送入 SoftMax 层, 该层生成类的概率分布。此外, 我们在每个 Conv、LSTM 和稠密层中引入高斯噪声, 以提

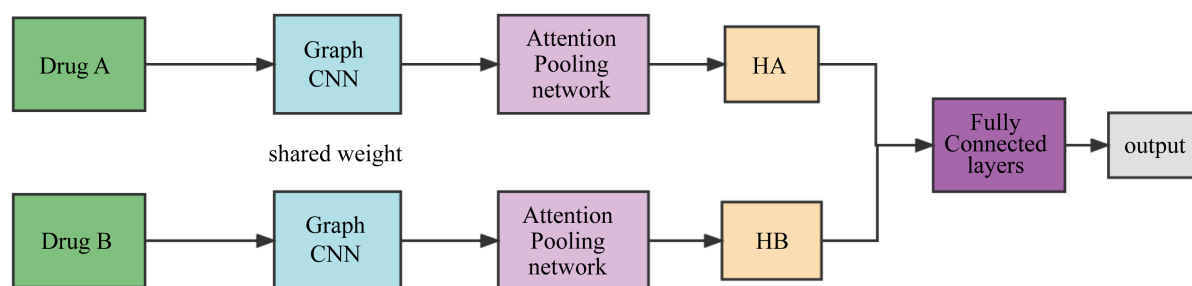


高模型的泛化能力。在 5 倍交叉验证测试中，分别获得了高达 0.94, 0.92, 0.80 的 AUPR, F1-score 和 MCC，显示在预测潜在的 DDI 方面有很高的可信度。具体流程如图 3 所示。



**Figure 3.** Flow chart of Conv-LSTM model  
**图 3.** Conv-LSTM 模型流程图

在 2019 年, Zhong 等人提出一个增大图形卷积神经网络模型预测 DDI 的 [11]。该方法将图卷积神经网络与注意力池化网络相结合, 提取药物对之间的结构关系并进行 DDI 预测。具体地说, 我们使用 RDKit 将药物 SMILES 字符串转换为图形拓扑。然后应用权重共享图卷积神经网络(GCNN)生成图表示。在这样的 GCNN 层之后, 我们嵌入了一个注意力池化网络来计算两种药物如何相互作用。在两个输入对之间建立一个紧密联系的双向注意力机制, 算法通过应用 CNN 或者 LSTM 去计算一个软结合  $G = \tanh(QTWA)$ , 表示两者之间的交互相关分数。最后通过一个 SoftMax 规范化操作, 应用行和列的最大池化, 被分隔的两边整合了共同的信息, 即图中的  $h_A$ ,  $h_B$ 。然后将  $h_A$  和  $h_B$  连接起来, 输入完全连接的层。最后, 一个简单的 Sigmoid 函数确定最终的分类。最后实验通过五折交叉验证显示, ROC 为 0.988, F1-score 为 0.956, AUPR 为 0.986。此外, 该模型还可以通过不同颜色的原子来揭示两种 DDI 药物在结构上的相互作用。这可能有助于药物发现过程中的药物设计。具体流程如图 4 所示。



**Figure 4.** Model flow chart  
**图 4.** 模型流程图

在 2015 年, Kc 等人提出了一种名为高阶图卷积网络(HOGCN)来聚合来自高阶邻域的信息, 去用于生物学实体对之间的交互预测[12]。具体来说, HOGCN 收集不同距离上邻居的特征表示, 并学习它们的线性混合, 以获得有用的生物学实体的代表。HOGCN 使用交互网络结构  $G$  或特征  $X$  去学习每个生物学实体的表示, 以交互网络  $G$  为输入, 重构交互网络中的边的高阶图卷积网络(HOGCN)。HOGCN 主要有两个组成部分: 编码器: 一种高阶图卷积编码器, 它在交互图  $G$  上运行, 并通过在不同距离上聚

集来自  $k$  阶邻域的特征来产生生物医学实体的表示。该层可以学习混合不同距离的邻居的特征表示，以进行交互预测。解码器：利用 HOGC 层产生的特征表示重建输入交互网络  $G$  中的边缘。采用双线性层融合生物医学实体  $v_i$  和  $v_j$ ，并且学习边  $e_{ij}$ 。边  $e_{ij}$  被输入到两层的全连接神经网络中去预测实体  $v_i$  和  $v_j$  相互作用的可能性。编码器 - 解码器的方法使 HOGCN 成为一个端到端可训练的交互模型。在药物 - 药物相互作用上表明，HOGCN 具有更准确和标准的预测。在考虑不同距离下邻居特征表示的情况下，HOGCN 在有噪声、稀疏的交互网络上表现良好。此外，通过基于此文献的案例研究，验证了一些新颖的交互预测。具体流程如图 5 所示。

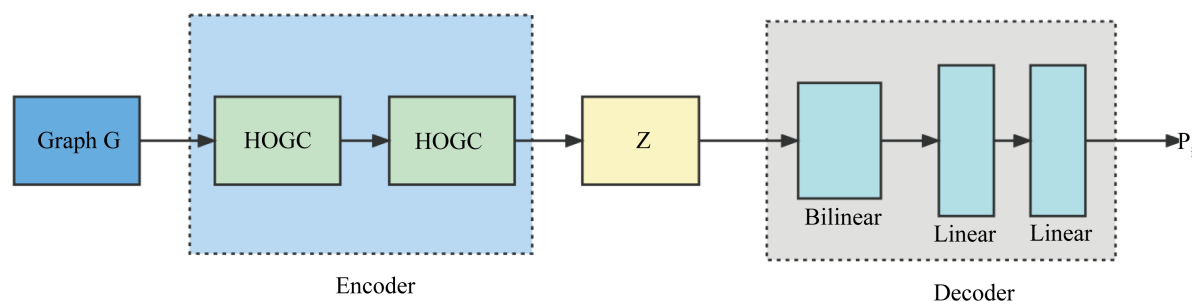


Figure 5. Flow chart of HOGCN model

图 5. HOGCN 模型流程图

## 5. 结果比较

表 2 为五种基于图卷积网络模型在药物相互作用关系预测的比较。

Table 2. Comparison of experimental results of five models

表 2. 五种模型实验表现结果对比

Year	Method	Database	Performance (AUPR)
2022	GCNMK	DrugBank	0.9508
2022	deepMDDIs	DrugBank	0.962
2019	Conv-LSTM	DrugBank、KEGG Drug、PharmGKB 和 OFFSIDES	0.94
2019	GCNN + Attentive pooling	DrugBank	0.986
2015	HOGCN	BioSNAP-DDIs	0.897 ± 0.003

## 6. 结论与展望

通过实验结果证明，基于图卷积网络预测药物之间的相互作用取得了有效的结果，证明了预测确实是有必要的，这样能节省大量的时间和物质成本。但即便如此，模型也存在局限性。在 GCNMK 模型中构造 DDIs 图和生成药物集时，实验中的药物至少有一个 DDIs。但这里将不含任何已知 DDIs 的药物移除。因此，模型不能在分离的药物中识别 DDIs。未来的研究有几个方向，在 GCNMK 的 DDIs 图中，边属于同一类型。我们可以将其应用于任何异质网络，比如药物疾病网络。药物 - 疾病关联的描述包括两种类型：疗法和标记物/机制，这可能有助于采用 GCN 模型。未来的另一个方向是区分更多类型的 DDIs。根据它们的功能，每种类型的 DDIs 都可以被用来构造一个图核，并且该模型有潜力来识别预测的 DDIs

的特定类型。在另外一个 HOGCN 模型中只考虑已知的交互来标记潜在的交互。生物医学信息的其他来源,如生物医学实体的各种物理化学和生物特性,可以提供关于交互作用的额外信息,可以计划研究将这些特征集成到模型中。由于 HOGCN 模型在不同的距离上聚集邻域信息,并可以标记新的交互作用,因此,以输入交互网络 G 的一个小子图的形式为预测提供可说明的解释将是有趣的。

在这篇综述中,我们全面回顾了近些年来使用图卷积网络来进行药物之间相互作用关系预测的文章,这些文章中的模型已经成功预测了药物之间的相互关联,并取得了优良结果。此外,在文章最后我们也总结了上面模型中存在的缺点,并为未来从事此研究方向的学者提出了一些建议,希望这些能帮助改进现有的方法来取得更优的结果。

## 参考文献

- [1] Lecun, Y., Bottou, I., Bengio, Y. and Hainier, P. (1998) Uradiant-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, o3y. <https://doi.org/10.1109/5.726791>
- [2] Zhou, F.-Y., Jin, L.-P. and Dong, J. (2017) Review of Convolutional Neural Network. *Chinese Journal of Computers*, **40**, 1229-1251.
- [3] Zhang, S., Uong, Y.-H. and Wang, J.J. (2019) The Development of Deep Convolution Neural Network and Its Applications on Computer Vision. *Chinese Journal of Computers*, **42**, 453-482.
- [4] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [5] Hu, B.T., Lu, Z.D., Li, H. and Chen, Q.C. (2014) Convolutional Neural Network Architectures for Matching Natural Language Sentences. *Proceedings of the Advances in Neural Information Processing Systems*, Montreal, 8-13 December 2014, 2042-2050.
- [6] Hinton, G., Deng, I., Yu, D., et al. (2012) Deep Neural Networks Ior Acoustic Modeling in Speech Recognition: The Shared Views of Lour Research Groups. *IEEE Signal Processing Magazine*, **29**, 82-97. <https://doi.org/10.1109/MSP.2012.2205597>
- [7] 王竟成, 张勇, 胡永利, 等. 基于图卷积网络的交通预测综述[J]. 北京工业大学学报, 2021, 47(8): 954-970.
- [8] Wang, F., Lei, X., et al. (2022) Predicting Drug-Drug Interactions by Graph Convolutional Network with Multi-Kernel. *Briefings in Bioinformatics*, **23**, bbab511. <https://doi.org/10.1093/bib/bbab511>
- [9] Feng, Y.-H., et al. (2022) deepMDDIs: A Deep Graph Convolutional Network Framework for Multi-Label Prediction of Drug-Drug Interactions. *Analytical Biochemistry*, **646**, Article ID: 114631.
- [10] Karim, M.R., Cochez, M., Jares, J.B., et al. (2019) Drug-Drug Interaction Prediction Based on Knowledge Graph Embedding and Convolutional-LSTM Network. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, Niagara Falls, 7-10 September 2019, 113-123. <https://doi.org/10.1145/3307339.3342161>
- [11] Zhong, Y. and Chen, X. (2019) Graph-Augmented Convolutional Networks on Drug-Drug Interactions Prediction.
- [12] Kc, K., Li, R., Cui, F., et al. (2015) Predicting Biomedical Interactions with Higher-Order Graph Convolutional Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **19**, 676-687. <https://doi.org/10.1109/TCBB.2021.3059415>