

## 完整文章

# AttentionMGT-DTA：一种使用图变换和注意力机制的多模态药物-靶点亲和力预测

Hongjie Wu<sup>a</sup>, Junkai Liu<sup>a,b</sup>, Tengsheng Jiang<sup>c</sup>, Quan Zou<sup>b</sup>, Shujie Qi<sup>a</sup>, Zhiming Cui<sup>a</sup>, Prayag Tiwari<sup>d,\*</sup>, Yijie Ding<sup>b,\*</sup>

<sup>a</sup>苏州科技大学电子信息工程学院, 中国苏州 215009

<sup>b</sup>长江三角洲地区研究院(衢州), 电子科技大学, 衢州, 324003, 中国

<sup>c</sup>南京医科大学古苏校区, 苏州, 215009, 中国

<sup>d</sup>瑞典哈尔姆斯理工大学信息技术学院

## 文章信息

关键词: 药物-靶点亲和力 图神经网络 图变换器 注意力机制 多模态学习

## 摘要

药物-靶点亲和力 (DTA) 的准确预测是药物发现和设计中的关键步骤。传统实验非常昂贵且耗时。最近, 深度学习方法在DTA预测中取得了显著的性能提升。然而, 基于深度学习的模型面临的一个挑战是药物和靶点的适当和准确表示, 尤其是缺乏对靶点表示的有效探索。另一个挑战是如何全面捕捉不同实例之间的交互信息, 这对于预测DTA也很重要。在本研究中, 我们提出了AttentionMGT-DTA, 一种用于DTA预测的多模态注意力模型。AttentionMGT-DTA分别通过分子图和结合口袋图来表示药物和靶点。采用了两种注意力机制来整合和交互不同蛋白质模态和药物-靶点对之间的信息。实验结果表明, 我们的模型在两个基准数据集上优于最先进的基准模型。此外, AttentionMGT-DTA通过建模药物原子与蛋白质残基之间的相互作用强度, 还具有很高的可解释性。我们的代码可在 <https://github.com/JK-Liu7/AttentionMGT-DTA> 获取。

## 1. 引言

药物发现是一个昂贵且耗时的过程。新药获批的典型过程通常需要28亿美元, 并需要10到15年的时间 (Wouters、McKee和Luyten, 2020; Yang、Ding、Tang和Guo, 2021)。然而, 大多数临床试验阶段的药物尚未获得批准并进入市场 (Newman和Cragg)。

2020年)。近来, 药物与靶蛋白之间相互作用的鉴定在药物发现中起着至关重要的作用, 这是一个研究热点, 并且已得到广泛研究 (Ezzat、Wu、Li和Kwoh, 2018年; Qian、Ding、Zou和Guo, 2022年)。许多传统方法已被用于预测给定的药物-靶点对的相互作用, 将其作为二元分类任务 (Bahi和Batouche, 2021年; Ding、Tang和Guo, )

(2021年)。然而, 结合亲和力是一个连续值, 反映了药物与靶点之间的结合强度 (丁、唐、郭和邹, 2022年)。因此, 预测药物-靶点亲和力 (DTA) 的回归任务也已成为药物发现和药物再定位领域的一个关键问题。

随着生物医学数据的大量应用 (丁、蒂瓦里、郭和邹, 2022) 以及计算资源的改善 (赵和全, 2021; 王、翟、丁和邹, 2023; 张、蒂瓦里等人, 2021), 基于深度学习的方法 (克利宁格和克洛克, 2021; 毛、施和周, 2021) 已在生物信息学中得到广泛应用, 尤其是在差热分析 (DTA) 预测领域 (吴、凌等人, 2021; 张、宋等人, 2021)。一般来说, 基于深度学习的模型包括数据预处理、药物特征提取模块、蛋白质特征提取模块和预测模块 (金伯、陈和沃尔卡默, 2021)。药物和蛋白质最广泛使用的一维 (1D) 序列表示分别是简化分子输入规范 (SMILES) 和氨基酸序列。对于基于深度学习的模型, 1D表示向量被输入到神经网络中以预测结合亲和力。例如, 奥兹图尔克、奥兹古尔和奥兹基林利 (2018) 使用卷积神经网络 (CNN) 作为药物和靶点的编码器。李、金和南 (2019) 使用分子指纹作为药物表示, 并使用线性层作为药物编码器。此外, 循环神经网络

## \*通讯作者

E-mail addresses: hongjiuwu@mail.usts.edu.cn (H. Wu), 1737969704@qq.com (J. Liu), 1911042002@post.usts.edu.cn (T. Jiang), zouquan@nclab.net (Q. Zou), 997196224@qq.com (S. Qi), zmcui@usts.edu.cn (Z. Cui), prayag.tiwari@ieee.org (P. Tiwari), wuxi\_dyj@csj.uestc.edu.cn (Y. Ding).

<https://doi.org/10.1016/j.neunet.2023.11.018>

Received 26 June 2023; Received in revised form 29 September 2023; Accepted 7 November 2023

Available online 11 November 2023

0893-6080/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(循环神经网络(RNN))和长短期记忆网络(LSTM)也被用作差热分析(DTA)预测中的特征提取器(卡里米、吴、王和沈, 2019年; 郑、李、陈、徐和杨, 2020年)。

尽管基于序列的方法在溶解热分析(DTA)预测中取得了显著的性能,但这并不是一种自然的药物表示方式,因为分子结构信息可能会丢失(Wang、Tang、Ding和Guo, 2021)。目前,分子图在DTA预测中也被广泛应用(Jiang等人, 2020; Li等人, 2020; Nguyen等人, 2020; Torng和Altman, 2019; Tsubaki、Tomii和Sese, 2018)。例如, Nguyen等人(2020)提出了GraphDTA,使用各种图神经网络(GNN)变体模型来提高预测效果。另一方面,基于图的蛋白质表示也被用于预测DTA(Jiang等人, 2020; Li、Zhang、Guan和Zhou, 2022a; Nguyen、Nguyen、Le和Tran, 2022a; Zheng等人, 2020)。Jiang等人(2020)提出了名为DGraphDTA的模型,利用从蛋白质序列获得的二维接触图作为蛋白质的描述符,以获取更多的结构信息。与基于一维序列的方法相比,二维基于图的模型可以利用药物和蛋白质更多的拓扑信息,这在DTA预测中显示出巨大的优势。

尽管在DTA预测方面取得了近期进展,但现有方法仍存在两个主要缺陷,限制了性能。大多数基于图的模型使用二维接触图或距离图来表示目标蛋白质(Jiang等, 2020; Li等, 2020; Zheng等, 2020)。然而,这些图表示方法只是对蛋白质结构的近似抽象,无法准确描述蛋白质的复杂三维(3D)结构(Ding、Tang和Guo, 2020a)。因此,结构意识的缺乏限制了所提出模型的准确性和泛化能力。为此,我们倾向于在我们的方法中使用基于三维结构的特征和基于一维序列的特征。一方面,AlphaFold2的发展为我们大规模使用蛋白质的三维结构提供了可能性。这些准确的三维结构包含了大量关于蛋白质结合口袋的实用性和构型的相关信息,对药物分子之间的结合过程有一定的积极影响。另一方面,一维序列表示可以作为三维结构的补充,提供上下文和生物学信息。我们认为,为蛋白质提供的额外的基于高阶序列的嵌入能够为我们的模型提供清晰且有效的知识,以区分不同的蛋白质,通过这种方式,DTA预测模型能够充分利用这些特征来衡量数据集中已见和未见蛋白质之间的关联,从而提高性能和泛化能力。

此外,在整合基于一维序列的特征和基于图的特征时,现有方法通常使用全局池化和简单连接来形成最终的表示(Wang、Zheng等人, 2022年; Wu、Gao、Zeng、Zhang和Li, 2022年)。此类操作在药物编码器和蛋白质编码器的表示的整合和融合中也非常常见,它们忽略了复杂的多模态交互作用。此外,简单的连接过程无法预测哪些蛋白质结合位点对与给定药物的结合贡献最大。因此,在我们的方法中利用了交叉注意力模块来融合和整合蛋白质信息的内模态信息,并使用联合注意力机制来交互蛋白质和药物特征的跨模态信息。

为了缓解上述问题,我们提出了一种新颖的多模态深度学习方法,名为AttentionMGT-DTA,用于DTA预测。首先,我们基于二维分子结构和三维蛋白质折叠分别构建了药物图和蛋白质口袋图,其中包含了节点和边的特征。然后,将药物图和蛋白质图输入到编码器中,以获得特征嵌入。药物编码器包含一个图变换器架构,而蛋白质编码器还包含一个基于一维序列嵌入的模态交互模块。最后,提取的特征向量被输入到基于注意力的预测模块中,以完成不同实例之间的信息融合并预测DTA。我们的主要贡献总结如下: · 我们的

模型是首次大规模应用AlphaFold2预测的蛋白质结构在DTA预测中的应用。我们基于AlphaFold数据库构建了一个残基级别的蛋白质口袋图来表示目标蛋白质。然后,分别使用两个图变换器对蛋白质口袋图和药物分子图进行特征提取。

- 为了丰富蛋白质的表示,我们引入了蛋白质氨基酸序列的自监督预训练嵌入。据我们所知,我们提出的模型是第一个使用交叉注意力来整合蛋白质两种模态(即1D序列和3D图)信息的模型。
- 我们引入了一种联合注意力机制,利用原子-残基相互作用矩阵生成亲和力结果。这使得我们的模型能够获得具有高度可解释性的交互式药物-靶点对嵌入,并了解蛋白质的哪些残基与药物原子相互作用。

本文的结构如下。在第二部分,介绍了与我们的研究相关的相关工作,包括DTA预测的发展以及蛋白质表示方法的研究。第三部分介绍了我们提出的AttentionMGT-DTA框架,其中包括药物分子图构建、蛋白质口袋图构建、图变换模型、蛋白质内模态交互以及药物-蛋白质跨模态交互的描述。第四部分展示了实验结果并进行了讨论。最后,第五部分给出了结论和未来的研究方向。

## 2. 相关工作

在本节中,我们从两个方面介绍相关工作。首先,我们简要回顾了现有的差热分析(DTA)预测的研究和技术。此外,还介绍了该领域中蛋白质表示方法的当前发展情况。

### 2.1. 差热分析(DTA)预测方法

#### 2.1.1. 基于深度学习的差热分析(DTA)预测方法

最近,深度学习技术发展迅速,并在药物-蛋白质相互作用(DTA)预测的计算方法中取得了巨大成功。目前,大多数基于深度学习的DTA预测模型直接将药物和蛋白质的表示作为特征,使用各种类型的深度学习模型对其进行编码,从中提取信息,并将各自的表示结合起来以预测结合亲和力(Dhakal、McKay、Tanner和Cheng, 2021; Ding、Tang和Guo, 2020b)。Ozturk等人(2018年)首次将卷积神经网络(CNN)应用于DTA预测。在他们提出的DeepDTA模型中,分别使用CNN提取药物和蛋白质的低维特征。将获得的特征向量输入到全连接层以计算结合亲和力。后来,他们添加了新的特征,包括蛋白质上的基序和结构域信息,并提出了一种名为WideDTA(Ozturk、Ozgirimli和Ozgur, 2019年)的模型,在两个基准数据集上比DeepDTA表现更好。Lee等人(2019年)提出了DeepConv-DTI,使用CNN对不同长度的氨基酸序列进行卷积,以获得蛋白质的局部残基模式。Rifaioğlu等人(2020年)首次使用二维图像结构来表示药物,减少了数据转换过程中的信息损失,并且所提出的DEEPScreen模型也取得了令人满意的性能。王、周、李和李(2021年)提出了DeepDTAF,它将蛋白质结合口袋作为输入特征,以整合局部和全局上下文信息。卡里米等人(2019年)提出了DeepAffinity,基于结合卷积神经网络(CNN)和循环神经网络(RNN)的seq2seq自编码器架构,考虑了残基或原子之间可能的依赖关系。李、赵和李(2022年)提出了一种CO-VAE方法,利用一种新颖的协同正则化变分自编码器来生成药物的简化分子线性输入规范(SMILES)字符串和目标序列,并采用协同正则化部分来获取结合亲和力。上述方法显示出模型具有很有前景的预测性能。

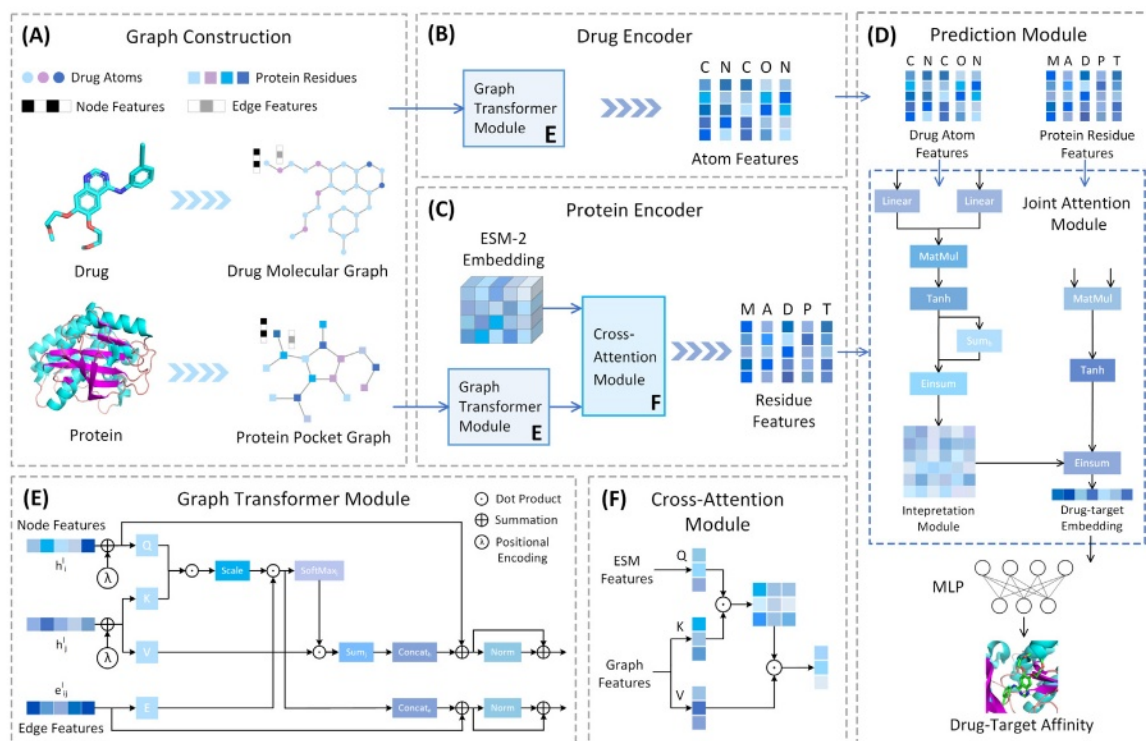


图1 AttentionMGT-DTA的整体架构。(A) 图构建模块, 在此模块中我们构建药物和蛋白质的图表示, 包括寻找蛋白质的结合位点。(B) 药物编码器模块, 通过图变换器提取药物特征。(C) 蛋白质编码器模块, 通过使用两种不同的信息模态提取蛋白质特征。(D) 预测模块, 用于预测药物-靶点对的结合亲和力, 其中包括一个解释模块, 用于深入解释哪些蛋白质残基与药物原子结合。(E) (B)和(C)中图变换器的网络结构。(F) (C)中详细的交叉注意力模块。

### 2.1.2. 基于图神经网络(GNN)的差热分析(DTA)预测方法

传统的卷积神经网络 (CNN) 和循环神经网络 (RNN) 模型将药物表示为数据格式的字符串, 结果可能会丢失分子的结构信息, 这可能会降低模型的预测能力和学习到的势空间的功能相关性 (Bagherian 等人, 2020 年)。药物和蛋白质可以自然地表示为具有原子级或残基级节点以及节点间边的图结构, 图神经网络 (GNN) 在考虑每个节点的邻节点的同时更新这些节点特征, 以提取全局结构特征 (Zhang 等人, 2022 年)。例如, Tsubaki 等人 (2018 年) 首次在差热分析 (DTA) 预测中应用 GNN, 分别使用 GNN 和 CNN 提取化合物和蛋白质的特征。Torng 和 Altman (2019 年) 将图卷积网络引入药物-蛋白质相互作用 (DTI) 预测中。在他们的方法中, 蛋白质结合口袋中的残基对应节点, 计算的特征向量表示它们的物理化学性质, 同时药物分子也被转换为图结构。Feng、Dueva、Cherkasov 和 Ester (2019 年) 提出了 PADME, 该方法结合了 ECFP 编码以及图表示, 并结合了蛋白质特征向量。由杨、钟、赵和陈宇谦 (2022 年) 提出的 MGraphDTA 将密集连接引入到图神经网络 (GNNs) 中, 并构建了一个超深的网络结构, 以同时捕获药物的局部和全局结构。对于目标蛋白质, 江等人 (2020 年) 首先提出了一种基于接触图的蛋白质图方法 DGraphDTA, 其中研究人员根据所获得的接触概率设定阈值 0.5 来确定残基节点是否相互连接, 所得结果对应于蛋白质的邻接矩阵。郑等人 (2020 年) 提出的研究使用了蛋白质的二维距离图作为输入, 并使用具有药物分子线性表示的视觉问答系统作为查询条件, 以获取所查询的药物和蛋白质是否相互作用的答案。受上述工作的启发, GEFA (阮、阮、勒和陈, 2022 年 b)、PSG-BAR (潘迪等人, 2022 年) 和 STAMP-DPI (王、郑等人, 2022 年) 也使用接触图来构建蛋白质图, 并添加了预训练的语言模型嵌入作为节点特征。

上述研究共同表明, 利用图数据结构能够有效地表示药物分子的结构和蛋白质残基。

### 2.1.3. 基于注意力机制的差热分析(DTA)预测方法

注意力机制在深度学习中逐渐变得愈发重要, 包括强大的自然语言处理 (NLP) 表示模型, 如 Transformer (Vaswani 等人, 2017 年) 和 BERT (Devlin、Chang、Lee 和 Toutanova, 2019 年), 并且在生物信息学中得到了广泛重视。AttentionDTA (Zhao、Xiao、Yang、Li 和 Wang, 2019 年) 首次将注意力机制与结合亲和力相关联, 利用注意力机制来考虑蛋白质中哪些子序列对药物更重要, 以及药物中哪些子序列对蛋白质更重要, 从而使模型更具表现力。Chen 等人 (2020 年) 提出了 TransformerCPI, 这是一种基于 Transformer 的模型, 利用 Transformer 解码器来提取基于序列模型的特定缺陷中的数据特征。Huang、Xiao、Glass 和 Sun (2020 年) 提出了 MolTrans, 这是一个通过两个 Transformer 模块从无标记数据中提取子结构之间相互作用的可解释模型。Huang 等人 (2022 年) 提出的 CoaDTI 也利用 Transformer 作为蛋白质序列的编码器, 以在氨基酸级别获得全局表示。李、张、关和周 (2022b) 通过引入互变型对变压器进行了改进, 其中使用了两个相互作用的变压器解码器来提取目标和药物的特征向量。ML-DTI (Yang、Zhong、Zhao 和 Chen, 2021) 设计了一种基于多头注意力的互学习机制, 以促进药物和蛋白质编码器之间的交互。由王、胡等人 (2022) 提出的 MGPII 利用变压器编码器来表示不同层次的特征, 完成了一个具有竞争性能的多粒度模型。赵、赵、郑和王 (2021) 提出了 HyperAttentionDTI, 采用注意力机制来对药物原子与蛋白质氨基酸之间的复杂分子间相互作用进行建模。



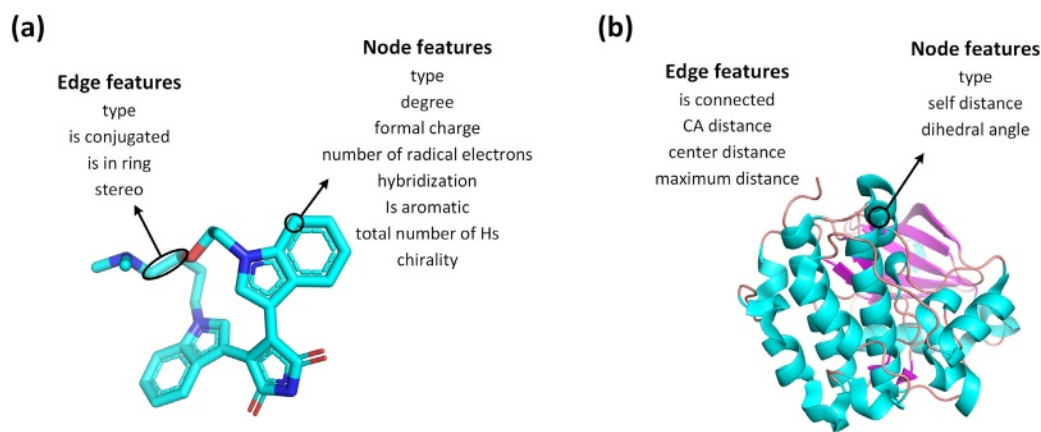


图 2. 药物和蛋白质的节点及边特征表示。(a) 原子级药物分子图，包含八个节点特征和四个边特征。(b) 残基级蛋白质口袋图，包含三个节点特征和四个边特征。

上述研究表明，基于图神经网络（GNN）的方法能够利用先进且精心设计的 GNN 模型，从药物分子和蛋白质残基中有效地提取拓扑信息。可以充分利用实体中包含的丰富的化学和生物信息来提高性能。此外，引入注意力机制在一定程度上能够提高可解释性和泛化能力，这也在这个研究领域引起了相当大的关注。

## 2.2. 蛋白质表示方法

蛋白质是参与各种生命活动的有机大分子，由不同的氨基酸序列组成，这些序列进而形成其独特的三维折叠结构，导致有序到无序以及构象的变化（丁、郭、蒂瓦里和邹，2023）。了解蛋白质的序列-结构-功能关系是蛋白质生物学中的核心问题，对于研究疾病机制、蛋白质设计和药物发现至关重要（贝普勒和伯杰，2021）。蛋白质数据表示的常用方法包括一维字符串、二维和三维图形结构。

蛋白质的完整序列，通常被称为蛋白质的初级结构，是指氨基酸在蛋白质中排列的顺序。蛋白质初级结构中的所有 20 种氨基酸都可以编码为单个字母，因此通常使用独热编码来表示蛋白质序列（ElAbd 等人，2020 年）。独热编码将表示氨基酸序列的字符转换为二进制向量，能够简短且高效地表示蛋白质的初级结构。基于上述技术，一些模型利用了将目标蛋白质表示为一维序列的 DTA 预测方法（Chen 等人，2020 年；Karimi 等人，2019 年；Nguyen 等人，2020 年；Ozturk 等人，2018 年）。

尽管序列是存储蛋白质一级结构信息的一种有效方式，但它无法提供蛋白质的三维结构信息。因此，蛋白质结构通常可以转化为具有化学性质的空间图，其中包含原子或残基节点和边。目前，广泛使用的蛋白质图表示包括二维和三维图。江等人（2020 年）首次提出将蛋白质的二维接触图作为其表示。每个蛋白质包含数百个氨基酸残基；然而，残基之间的连接只是一个长链，不包含任何空间信息。接触图是蛋白质结构的一种表示，是蛋白质三维结构的二维表示，也可以作为蛋白质结构预测的输出。在这种方法中，设定阈值为 0.5，以获得  $L \times L$  的接触图，其中  $L$  是节点（残基）的数量。另一方面，蛋白质数据库（PDB）是蛋白质数据的主要来源，但存在一个主要问题，即具有结构特征的蛋白质数量远远少于具有已测定氮

基酸序列的蛋白质数量（Li 等人，2020 年），这限制了利用蛋白质结构信息进行药物发现（Zhang 等人，2023 年）。因此，只有少数方法探索并开发了基于三维结构的蛋白质图表示方法（Wang、Liu 等人，2023 年；Wang、Zhang 等人，2023 年；Yazdani-Jahromi 等人，2022 年）。在本研究中，我们引入了一种基于蛋白质三维结构的蛋白质表示方法，大规模利用和开发由 AlphaFold2 预测的蛋白质结构，据我们所知，这是首次将这种方法应用于差热分析（DTA）预测问题，其最终动机是找到一种有效且准确的蛋白质表示方法并提高性能。

## 3. 材料与方法

### 3.1. 图形构建

#### 3.1.1. 药物图谱表示

药物分子被表示为一个图，以获取更多的化学和拓扑信息。药物的二维无向图可以描述为  $G_D = (V_D, E_D)$ ，其中节点和边分别表示药物原子与共价化学键。在药物图中， $V_D$  是具有特征向量的原子节点集合， $E_D$  表示为特征向量的边集合。图 2 展示了在我们的方法中构建药物图的节点和边特征，包括八种节点特征和四种边特征，这些特征在先前的研究中已被应用于表示药物分子（Jiang 等，2021；Wu, Jiang 等，2021）。

#### 3.1.2. 蛋白质图谱表示法

蛋白质的三维结构在药物结合中起着至关重要的作用。然而，部分蛋白质的晶体结构在 PDB 中不可用，这给利用这种结构带来了困难。更具体地说，数据集中可用的和不可用的蛋白质 PDB 结构数量可以在表 3 中找到。考虑到 AlphaFold2（Jumper, Evans, Pritzel, Green, & Hassabis, 2021）预测的三维蛋白质结构的精度和准确性，我们采用了基于 AlphaFold2 的高质量三维结构来替代 PDB 结构。在本研究中，我们通过 UniProt ID 从 AlphaFold 数据库（Varadi 等，2021）下载了蛋白质结构文件，这些文件都是 PDB 文件格式的。我们证明了 AlphaFold2 预测的高质量蛋白质结构也能有效地表示目标蛋白质，并提高

表1  
用于药物图表示的节点和边的特征。

类型	特征	描述	尺寸
节点特性	原子类型	“C” (碳)、 “N” (氮)、 “O” (氧)、 “F” (氟)、 “P” (磷)、 “S” (硫)、 “Cl” (氯)、 “Br” (溴)、 “I” (碘)、 “B” (硼)、 “Si” (硅)、 “Fe” (铁)、 “Zn” (锌)、 “Cu” (铜)、 “Mn” (锰)、 “Mo” (钼)、 “Ag” (银)、 “Au” (金)、 “Hg” (汞)、 “Pt” (铂)、 “Pd” (钯)、 “Ni” (镍)、 “Co” (钴)、 “Cr” (铬)、 “Mg” (镁)、 “Ca” (钙)、 “Na” (钠)、 “K” (钾)、 “Li” (锂)、 “Ba” (钡)、 “Sr” (锶)、 “Pb” (铅)、 “Sn” (锡)、 “V” (钒)、 “Ti” (钛)、 “Al” (铝)、 “Ga” (镓)、 “In” (铟)、 “Zr” (锆)、 “Hf” (铪)、 “Nb” (铌)、 “Ta” (钽)、 “Bi” (铋)、 “Sb” (锑)、 “As” (砷)、 “Se” (硒)、 “Te” (碲)、 “Ge” (锗)、 “Si” (硅)、 “P” (磷)、 “S” (硫)、 “Cl” (氯)、 “Br” (溴)、 “I” (碘)、 “F” (氟)、 “O” (氧)、 “N” (氮)、 “C” (碳) (独热编码)	17
	原子度	0 (铜) 2、3其他5 (独热编码)	7
	原子形式电荷	0 或者 1	1
	原子数 自由基电子数	0 或者 1	1
	原子杂化	“sp”、 “sp2”、 “sp3”、 “sp3d”、 “sp3d2”、 “其他” (独热编码)	6
	原子是芳香族的	0 或者 1	1
	原子总数 H	0、1、2、3、4 (独热编码)	5
边缘特性	原子的手性	“R”、 “S”、 “其他” (独热编码)	3
	粘键类型	“单键”、 “双键”、 “三键”、 “芳香键” (独热编码)	4
	键是共轭的	0 或者 1	1
	邦德在戒指里	0 或者 1	1
	邦德立体声	“STEREONONE”、 “STEREOANY”、 “STEREOZ”、 “STEREOE” (独热编码)	4

表 2  
蛋白质图表示的节点和边特征。

类型	特征	描述	尺寸
节点特性	残留类型	“G”、 “A”、 “V”、 “L”、 “I”、 “P”、 “F”、 “Y”、 “W”、 “S”、 “T”、 “C”、 “M”、 “N”、 “Q”、 “D”、 “E”、 “K”、 “R”、 “H”、 “金属”、 “其他” (独热编码)	22
	残余自距离	一个残基中所有原子的缩放距离的最大值和最小值。CA 原子与 O 原子之间、O 原子与 N 原子之间、C 原子与 N 原子之间的缩放距离	5
边缘特性	残余二面角	“φ”、 “ψ”、 “ω”、 “χ1”	4
	残留物已连接	0 或 1 的 CA 原子之间的标度距离 标度距离的中	1
	残余 CA 距离	心最大值和最小值之间的标度距离	1
	残余中心距离		1
	残余最大距离		2

DTA 预测性能使我们的模型优于几个基准模型。

随后，引入了萨贝里·法蒂和图辛斯基（2014 年）提出的方法来识别蛋白质的结合位点。通过该算法计算出的包围盒坐标可用于确定完整蛋白质结构中结合位点的部分，这有助于我们的模型减小输入数据规模并节省计算资源。此外，尽管这些确定的位点并不总是与确切的结合位点相对应，但使用蛋白质口袋的蛋白质图表示可以通过引导我们的模型专注于口袋中的通用拓扑特征，促进我们的方法具有高度的可泛化性。这一特性对于合理地推广到那些在训练数据中未曾见过的新蛋白质的情况至关重要。然后构建了蛋白质口袋图  $G_P=(V_P, E_P)$ ，其中节点表示蛋白质残基，边表示两个残基之间的相互作用。在此，我们将阈值设定为 10.0 埃，这意味着最小距离小于 10.0 埃的残基对被视为由边连接。阈值参数是经验性的，可优化的。考虑到模型的复杂性和计算资源的消耗，我们在残基级别而不是原子级别构建了蛋白质口袋图。对于图特征，我们根据沈等人（2022）的研究选择了三个节点特征和四个边特征，如图2所示。表1和表2详细列出了药物和蛋白质特征的列表。我们主要通过MDAnalysis包（Michaud-Agrawal, Denning, Woolf, & Beckstein, 2011）生成蛋白质特征。

3.2. 药物编码器模块

在我们的模型中，我们使用图变换器（Dwivedi & Bresson, 2020）框架作为药物编码器来提取药物图的节点表示，如图1所示。与其他G

NN模型相比，高级图变换模型具有以下优势：首先，带有边特征的图变换器可以在DTA预测中引入更明确的化学和生物特性作为边特征，即药物的共价化学键和蛋白质的残基相互作用。其次，图变换模型的架构使我们的方法能够捕获局部和全局信息，包括节点和连接信息的通用和特定模式。第三，该模型中创新的和改进的注意力机制有助于我们的方法测量每个原子和残基的重要性，这有助于识别确切的结合位点。此外，该模型中的位置编码对于编码节点位置信息和表示图拓扑结构至关重要，这是其他GNN模型所没有的属性。

对于具有节点特征 $a_i \in \mathbb{R}^{D_v \times 1}$ （对于节点I）和边特征 $\beta_{ij} \in \mathbb{R}^{D_E \times 1}$ （对于节点I和节点J之间的边）的药物图GD，首先将输入节点和边特征输入到一个线性投影层，以获得如下的隐藏表示 $h_i$ 和 $e_{ij}$ ：

$$h_i^0 = W_A^0 \alpha_i + b_A^0 \tag{1}$$

$$e_{ij}^0 = W_B^0 \alpha_i + b_B^0 \tag{2}$$

其中， $W_A0 \in \mathbb{R}^{D \times D_N}$  和  $W_B0 \in \mathbb{R}^{D \times D_E}$  是可学习的权重参数， $B_{0A}, B_{0B} \in \mathbb{R}^D$  是线性层的偏差。然后，位置编码被添加到节点特征中。

$$\lambda_i^0 = W_C^0 \lambda_i + b_C^0 \tag{3}$$

$$\hat{h}_i^l = h_i^0 + \lambda_i^0 \tag{4}$$

在我们的工作中，我们在图变换模型中利用拉普拉斯特征向量作为位置编码（Dwivedi 等人，2020 年）。特别是，每个图的拉普拉斯特征向量是通过图拉普拉斯矩阵的分解预先计算得出的，如下所示：

$$\Delta = I - D^{-1/2} A D^{-1/2} = U^T A U \tag{5}$$

其中,  $A$  是邻接矩阵,  $D$  是相应的图的度矩阵, 而  $\Lambda, U$  分别表示特征值和特征向量。节点  $i$  的  $K$  个小于平凡特征向量被用作其相应的位置编码, 表示为  $\lambda_i$ 。

图变换器更新节点和边的特征主要基于多头注意力机制。以下公式定义了第  $L$  层的详细更新过程:

$$Q_{ij}^{k,l} = Q^{k,l} \text{Norm}(h_i^l), K_{ij}^{k,l} = K^{k,l} \text{Norm}(h_j^l), V_{ij}^{k,l} = V^{k,l} \text{Norm}(h_j^l) \quad (6)$$

$$E_{ij}^{k,l} = E^{k,l} \text{Norm}(e_{ij}^l) \quad (7)$$

$$w_{ij}^{k,l} = \text{softmax}_j \left( \left( \frac{Q_{ij}^{k,l} h_i^l \cdot K_{ij}^{k,l} h_j^l}{\sqrt{d_k}} \right) \cdot E_{ij}^{k,l} e_{ij}^l \right) \quad (8)$$

$$\hat{h}_i^{l+1} = h_i^l + O_h \left( \text{Concat}_{k=1}^{h_{gt}} \left( \sum_{j \in N_i} w_{ij}^{k,l} V_{ij}^{k,l} h_j^l \right) \right) \quad (9)$$

$$\hat{e}_{ij}^{l+1} = e_{ij}^l + O_e \left( \text{Concat}_{k=1}^{h_{gt}} \left( w_{ij}^{k,l} \right) \right) \quad (10)$$

其中

$$Q = W_Q X_{ps}, K = W_K X_{pg}, V = W_V X_{pg} \quad (13)$$

以下所示, 被输入到具有残差连接和批量归一化层的正向网络中:

$$h_i^{l+1} = \hat{h}_i^{l+1} + W_{h2}^l \left( \text{ReLU} \left( W_{h1}^l \text{Norm}(\hat{h}_i^{l+1}) \right) \right) \quad (11)$$

$$e_{ij}^{l+1} = \hat{e}_{ij}^{l+1} + W_{e2}^l \left( \text{ReLU} \left( W_{e1}^l \text{Norm}(\hat{e}_{ij}^{l+1}) \right) \right) \quad (12)$$

在哪里  $W_{h1}, W_{h2} \in \mathbb{R}^{2D \times D}$  且  $W_{e1}, W_{e2} \in \mathbb{R}^{D \times 2D}$ 。通过上述方法 - E1 和 E2

图变换器模块, 即获得的节点特征, 即药物的原子特征  $X_D$  were, 用于进一步预测。

### 3.3. 蛋白质编码模块

同样, 图变换模型被用于对蛋白质图进行编码, 以学习残基特征。除了蛋白质口袋图的三维生物结构信息外, 我们的模型还使用了来自蛋白质序列的预训练嵌入来丰富多模态蛋白质信息表示。更具体地说, 我们引入了 ESM-2 (Lin 等人, 2023 年), 这是一种用于从单个序列预测蛋白质结构、功能和其他属性的蛋白质语言模型。预训练的 ESM-2 模型被用于将氨基酸序列编码为嵌入向量。每个嵌入向量都包含了来自一维蛋白质序列的大量上下文信息。此外, 我们获得了仅表示其结合口袋部分的蛋白质图嵌入, 而 ESM-2 嵌入表示的是整个蛋白质序列。为了解决这种不一致性, 我们对 ESM-2 嵌入进行了进一步处理。具体来说, 我们切割这些嵌入, 并只取其中对应于结合口袋的部分, 记为  $X_{ps} \in \mathbb{R}^{N^p \times D^p}$  (其中  $D_{ps}$  是蛋白质嵌入的维度)。我们认为, 仅使用不完整的预训练嵌入并不影响其效用, 因为由于上下文的性质, 嵌入的片段也包含局部特征信息。

连接策略是一种简单的整合方法, 通过编码和混合两种模态的嵌入来保留来自单个模态的信息, 在先前的研究中广泛应用 (Wang, Zheng, 等, 2022; Xu, Hu, Leskovec, & Jegelka, 2019)。然而, 不同的蛋白质嵌入模态是相互关联的, 简单的连接不足以捕捉交互信息并揭示它们之间的关系, 一些基本的结构或序列信息可能会因此丢失。因此, 我们引入了一个交叉注意力模块来执行蛋白质的一维序列和三维图之间的模态交互,

如图1所示。引入交叉注意力机制有以下好处: 一方面, 通过注意力分数来衡量结合过程中每个残基的重要性, 可以获得全面合理的特征嵌入。另一方面, 该模块可以融合和整合来自双模态的大量生物学特性, 使我们的模型学习到更多信息特征表示。在这个模块中, 给定来自图变换器模块的嵌入  $X_{pg} \in \mathbb{R}^{N^p \times D^p}$  以及来自预训练模型的嵌入  $X_{ps} \in \mathbb{R}^{N^p \times D^p}$ , 我们计算多模态输出  $X_p$  as 如下:

$$Q = W_Q X_{ps}, K = W_K X_{pg}, V = W_V X_{pg} \quad (13)$$

$$X_p = \text{attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (14)$$

交叉注意力模块通过两种编码的蛋白质嵌入的模态交互, 促进我们的模型学习独立蛋白质模态 (1D 序列和 3D 结构) 之间的关系。因此, 我们的模型能够从相关的蛋白质模态中捕获更全面和有效的信息, 从而促进了 DTA 预测。

### 3.4. 预测模块

在药物和蛋白质编码器之后, 将学习到的特征嵌入向量  $X_D$  and 药物  $X_p$  were 蛋白质 输入到预测网络中, 以输出药物 - 靶点交互的嵌入  $X_{DP}$ 。与之前的研究中在连接这两个表示之后使用多层感知机 (MLP) 不同, 我们采用了联合注意力机制 (Karimi 等人, 2019 年; Karimi, Wu, Wang 和 Shen, 2021 年) 来整合和交互信息, 如图 1 所示。药物嵌入  $X_D \in \mathbb{R}^{N^D \times D^D}$  和蛋白质嵌入  $X_P \in \mathbb{R}^{N^P \times D^P}$  的处理方式如下, 以获得注意力分数:

$$N_{ij} = \tanh \left( (x_i^D)^T W_{A1} x_j^P \right) \quad \forall i = 1, \dots, n_d, \quad \forall j = 1, \dots, n_p \quad (15)$$

$$\alpha_{ij} = \frac{\exp(N_{ij})}{\sum_{i', j'} \exp(N_{i' j'})} \quad \forall i = 1, \dots, n_d, \quad \forall j = 1, \dots, n_p \quad (16)$$

其中,  $I$  是药物原子的索引,  $J$  是蛋白质残基的索引,  $W_{A1}$  表示权重矩阵。注意力得分  $\alpha_{ij}$  构成注意力矩阵  $A \in \mathbb{R}^{DD \times DP}$ , 其中每个位置表示第  $i$  个药物原子和第  $j$  个蛋白质残基之间的相互作用强度。然后, 药物 - 靶点表示通过  $X_{DP}$  进行嵌入:

$$f_{ij} = \tanh \left( W_{A2} x_i^D + W_{A3} x_j^P + b_A \right) \quad (17)$$

$$X_{dp} = \sum_{i,j} f_{ij} \alpha_{ij} \quad (18)$$

其中,  $W_{A2}$ ,  $W_{A3}$  和  $b_A$  是可学习的参数。

然后, 计算得出的药物-靶点相互作用特征被传递给多层感知机 (MLP) 以获得最终的亲和力值。在本研究中, MLP 由三个具有 ReLU 激活函数和随机失活层的全连接层组成。由于 DTA 预测是一项回归任务, 均方误差 (MSE) 被用作损失函数:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (P_i - Y_i)^2 \quad (19)$$

其中,  $N$  表示药物 - 靶点对的数量,  $P_i$  and  $Y_i$  are 表示预测的结合亲和力值和真实值。

表3  
精炼基准数据集摘要

数据集	毒品	蛋白质	可用的 PDB 数据库	不可用的 PDB 数据库	交互	活动	未激活
戴维斯	68	361	275	86	24 548	1649	22 899
基巴	2052	229	194	35	117 148	24 543	九十二百四十一

表4  
AttentionMGT-DTA 的超参数设置

超参数	设置
蛋白质残基图的阈值	[5, 8, 10, 12, 15]
图变换器的层数	[2, 3, 5, 10]
注意力头的数量	[1, 2, 4, 8]
图变换器的辍学率	0.2
药物嵌入的尺寸	[32, 64, 128, 256]
蛋白质嵌入的维度	[32, 64, 128, 256]
学习率	1 乘以 10 的负 4 次方
批量大小	60
纪年; 时代	1000

表5  
在戴维斯数据集上，我们的模型与基准方法之间的性能比较。

数据集	模型	R2M	持续集成	均方误差
戴维斯	深度差热分析 (DeepDTA)	0.690 (0.035)	0.882 (0.016)	<b>0.191</b>
	注意差值分析 (DTA)	0.697 (0.005)	0.888 (0.007)	0.195
	图形差热分析 (DTA)	0.682 (0.028)	0.876 (0.019)	0.194
	变压器 CPI	0.658 (0.033)	0.872 (0.015)	0.201
	机器学习 - 确定性时间序列分析	0.627 (0.022)	0.869 (0.009)	0.196
	最小保留价格 (MGPLI)	0.620 (0.017)	0.884 (0.004)	0.218
	注意: MGT - DTA	<b>0.699 (0.027)</b>	<b>0.891 (0.005)</b>	0.193

## 4. 实验与结果

### 4.1. 数据集

我们在两个基准数据集，即戴维斯数据集和KIBA数据集（戴维斯等人，2011年；唐等人，2014年）上，将我们的AttentionMGT-DTA模型与其他基准模型进行了比较。结合亲和力和值通常通过诸如解离常数（ $K_D$ ）、抑制常数（ $K_I$ ）和半抑制浓度（ $IC_{50}$ ）等指标来表示。在戴维斯数据集中，亲和力和值通过 $K_D$ 来评估，该指标反映了激酶蛋白家族及其相关抑制剂的解离常数的恒定值的选择性测量。对于KIBA数据集，亲和力和值通过一种名为KIBA的方法进行测量，该方法利用 $IC_{50}$ 中包含的统计信息来优化它们之间的一致性。KIBA数据集中的亲和力和值主要在10到13之间，大多数在11左右。此外，原始数据集中的重复样本被去除，以减少它们对模型训练的影响。表3显示了经过改进的基准数据集的汇总统计数据。

### 4.2. 实验设置

在我们的实验中，AttentionMGT-DTA 是通过 Pytorch 实现的。使用了 Adam 优化器（Kingma & Ba, 2017）进行模型训练，学习率为 0.0001。还采用了学习率衰减策略，即在 50 个周期中，当测试数据集上的均方误差指数没有改善时，学习率降低 20%。我们使用 Nvidia RTX 3090 GPU 进行实验。超参数优化的最佳设置如表 4 所示。

### 4.3. 评估指标

对于差热分析 (DTA) 预测的回归任务，我们使用均方误差 (MSE)、一致性指数 (CI) 和向均值回归 ( $R^2_M$ ) 来评估我们模型的性能。一致性指数 (CI, Gonen 和 Heller, 2005) 是一种评估

表6  
在 KIBA 数据集上，我们的模型与基准方法之间的性能比较。

数据集	模型	R2M	持续集成	均方误差
基巴	深度差热分析 (DeepDTA)	0.766 (0.085)	0.892 (0.026)	0.152
	注意差值分析 (DTA)	0.742 (0.015)	0.880 (0.001)	0.158
	图形差热分析 (DTA)	0.760 (0.049)	0.888 (0.023)	0.203
	变压器 CPI	0.721 (0.022)	0.875 (0.009)	0.205
	机器学习 - 确定性时间序列分析	0.764 (0.025)	0.890 (0.005)	0.177
	最小保留价格 (MGPLI)	0.753 (0.016)	0.891 (0.004)	0.159
	注意: MGT - DTA	<b>0.786 (0.018)</b>	<b>0.893 (0.001)</b>	<b>0.140</b>

通过计算预测值与实际值之间的差异来反映结果正确性的度量标准，如下所示：

$$CI = \frac{1}{Z} \sum_{\delta_j > \delta_i} h(b_i - b_j) \quad (20)$$

$$h(x) = \begin{cases} 0 & x < 0 \\ 0.5 & x = 0 \\ 1 & x > 0 \end{cases} \quad (21)$$

均方误差 (MSE) 是衡量误差的一个常用指标。给定 N 个样本，以及相应的预测亲和值  $\hat{Y}_i$  和真实亲和值  $Y_i$ ，均方误差的定义如下：

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (22)$$

$R^2_M$  是一种评估外部预测性能的度量标准。当且仅当  $R^2_M \geq 0.5$  时，一个模型才被视为可接受的。 $R^2_M$  定义如下：

$$r_m^2 = r^2 * \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (23)$$

其中， $R$  表示观测值和预测值之间（含截距）的相关系数的平方， $R_0$  表示不含截距时的相关系数。

### 4.4. 结果

#### 4.4.1. 在基准数据集上的性能

在此，我们将我们的模型与以下基准进行了比较：Deep-DTA（Ozturk 等人，2018 年）、AttentionDTA（Zhao 等人，2019 年）、GraphDTA（Nguyen 等人，2020 年）、TransformerCPI（Chen 等人，2020 年）、ML-DTI（Yang、Zhong 等人，2021 年）和 MGPLI（Wang、Hu 等人，2022 年），这些都是最先进的方法。

DeepDTA（Ozturk 等人，2018 年）分别采用了两个卷积神经网络（CNN）模块从药物的简化分子线性输入规范（SMILES）和蛋白质序列中提取嵌入。然后，使用了一个多层感知机（MLP）从连接后的特征进行预测。

AttentionDTA（赵等人，2019 年）引入注意力机制来衡量药物和蛋白质中不同子序列的重要性，这增强了其表示能力。

GraphDTA（Nguyen 等人，2020 年）是一种基于图神经网络（GNN）的方法，它利用各种类型的 GNN 模型对药物分子图进行编码，并利用卷积神经网络（CNN）对蛋白质序列进行编码。

TransformerCPI（Chen 等人，2020 年）提出了一种改进的 Transformer 架构，以完成药物和蛋白质序列表示之间的信息交互。



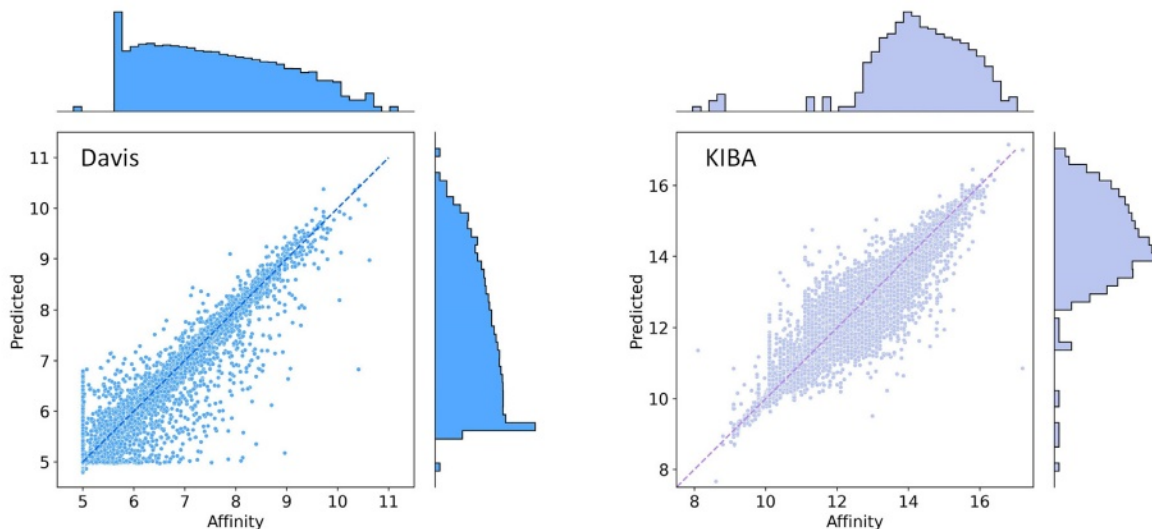


图 3. 在 Davis 和 KIBA 数据集中，药物 - 靶点对的真实亲和力 (x 轴) 与预测亲和力 (y 轴)。

表7  
在冷启动设置下，我们的模型与基准方法在戴维斯数据集上的性能比较。

设置	模型	持续集成	均方误差	R2M
药物冷启动	深度差热分析 (DeepDTA)	0.633 (0.030)	0.675 (0.262)	0.062 (0.029)
	注意差值分析 (DTA)	0.649 (0.031)	<b>0.630 (0.274)</b>	0.091 (0.031)
	图形差热分析 (DTA)	0.660 (0.035)	0.722 (0.271)	0.108 (0.070)
	变压器 CPI	0.608 (0.059)	0.788 (0.255)	0.057 (0.044)
	机器学习 - 确定性时间序列分析	0.640 (0.030)	0.735 (0.195)	0.113 (0.068)
	最小保留价格 (MGPLI)	0.644 (0.051)	0.711 (0.189)	0.088 (0.045)
	电偶差热分析法	0.659 (0.055)	0.667 (0.129)	0.094 (0.069)
	注意: MGT - DTA	<b>0.696 (0.034)</b>	0.729 (0.213)	<b>0.162 (0.081)</b>
目标冷启动	深度差热分析 (DeepDTA)	0.780 (0.008)	0.424 (0.039)	<b>0.351 (0.036)</b>
	注意差值分析 (DTA)	0.748 (0.008)	0.490 (0.056)	0.246 (0.022)
	图形差热分析 (DTA)	0.755 (0.006)	0.494 (0.061)	0.277 (0.020)
	变压器 CPI	0.725 (0.009)	0.503 (0.051)	0.244 (0.026)
	机器学习 - 确定性时间序列分析	0.732 (0.008)	0.459 (0.032)	0.281 (0.025)
	最小保留价格 (MGPLI)	0.766 (0.005)	0.485 (0.048)	0.308 (0.019)
	电偶差热分析法	0.804 (0.006)	0.435 (0.042)	0.318 (0.018)
	注意: MGT - DTA	<b>0.829 (0.005)</b>	<b>0.422 (0.031)</b>	0.284 (0.017)
药物靶点的冷启动	深度差热分析 (DeepDTA)	0.597 (0.034)	0.679 (0.101)	0.037 (0.029)
	注意差值分析 (DTA)	0.560 (0.036)	0.676 (0.147)	0.017 (0.015)
	图形差热分析 (DTA)	0.603 (0.032)	0.779 (0.084)	0.058 (0.045)
	变压器 CPI	0.564 (0.029)	0.739 (0.155)	0.047 (0.042)
	机器学习 - 确定性时间序列分析	0.601 (0.033)	0.728 (0.143)	0.060 (0.033)
	最小保留价格 (MGPLI)	0.587 (0.027)	0.761 (0.148)	0.050 (0.052)
	电偶差热分析法	0.605 (0.026)	0.617 (0.105)	0.054 (0.045)
	注意: MGT - DTA	<b>0.613 (0.031)</b>	<b>0.612 (0.082)</b>	<b>0.065 (0.046)</b>

☑ML-DTI (Yang、Zhong 等人, 2021 年) 建立了一种相互学习的机制，从全局视角弥合特征提取模块之间的差距，提高了泛化和解释能力。

☑MGPLI (Wang、Hu 等人, 2022 年) 采用了变压器编码器来捕获原子与残基之间的潜在相互作用，并使用卷积神经网络 (CNN) 层来融合和整合多粒度信息。

值得一提的是，在基准实验中应用了 5 折交叉验证，以防止过拟合并确保比较的公平性。结果分别在表 5 和表 6 中呈现。

如图所示，AttentionMGT-DTA 在 Davis 数据集上的  $R^2_{metric}$  性能最佳，比基准模型提高了 0.2%。对于大规模数据集，即 KIBA 数据集，AttentionMGT-DTA 在所有指标方面都优于其他基准模型。具体而言，与最先进的方法相比，我们的模型将 CI 指标提高了 1.0%，将 MSE 指标降低了 1.2%。此外，我们的模型在  $R$

$^2_{Mindex}$  方面比以往方法提高了 2.0%。对于使用 GNN 作为药物编码器的 GraphDTA，AttentionMGT-DTA 在 CI、MSE 和  $R^2_M$  方面的平均改进分别为 1.0%、5.1% 和 2.2%。与基于 Transformer 的方法，即 TransformerCPI 和 MGPLI 相比，我们的方法在 DTA 预测方面也展示了其优势。这些改进表明了 3D 蛋白质图优越性，这使得我们的模型能够精确且全面地表示目标蛋白质。这些结果表明，AttentionMGT-DTA 在大规模数据集上通过基于注意力的多模态信息融合实现了极具竞争力的 DTA 预测性能。在提取目标蛋白质的特征时，序列和图结构信息的模型及相互作用显著改善了蛋白质表示。在提取药物和靶点的交互特征时，注意力模块能够保留药物 - 靶点对之间的综合信息。我们的模型在 Davis 和 KIBA 数据集上性能存在差异的可能原因是依赖于来自 AlphaFold 数据库的蛋白质三维结构。从 AlphaFold2 获得的蛋白质结构质量参差不齐，导致我们的模型在 Davis 数据集上的表现相对平庸。



表8  
在戴维森数据集上对蛋白质残基图阈值的参数设置分析。

阈值设定 (A)	边的平均数量	平均度数	R2M	持续集成	均方误差
5.0	6688	8.98	0.661 (0.023)	0.872 (0.007)	0.209 (0.006)
8.0	14 475	19.56	0.692 (0.029)	0.885 (0.006)	0.194 (0.009)
10.0	21564	29.19	<b>0.699 (0.027)</b>	<b>0.891 (0.005)</b>	<b>0.193 (0.010)</b>
12.0	30 756	41.66	0.694 (0.025)	0.887 (0.005)	0.195 (0.012)
15.0	47137	63.80	0.670 (0.031)	0.871 (0.006)	0.208 (0.011)

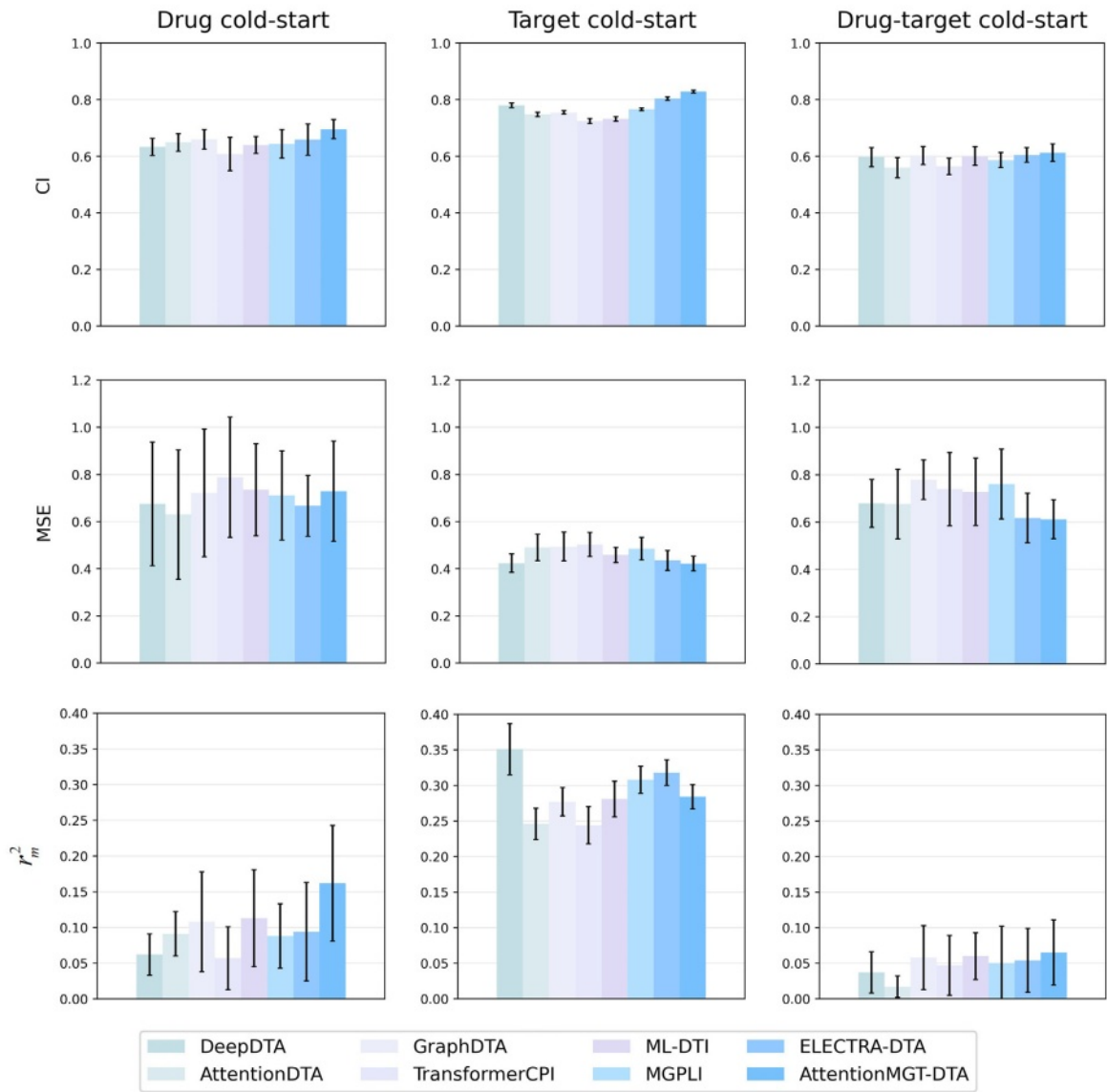


图 4. 在冷启动设置下，AttentionMGT-DTA 与基准方法在戴维森数据集上的性能比较。

为了进一步分析实验结果，我们还绘制了Davis和KIBA数据集的预测亲和力与真实亲和力。图3展示了这两个数据集的预测亲和力与实际亲和力的散点图。将x轴设置为真实亲和力，y轴设置为预测值，理想模型可以生成一条直线 $y = x$ 。如图3所示，样本点位于直线或接近直线，呈对称分布。此外，结果表明我们的模型在KIBA数据集上的表现更好，因为点高度密集地分布在直线 $y = x$ 周围。

4.4.2. 冷启动的性能

在药物-蛋白质相互作用（DTA）预测方法中，泛化能力和稳健性是关键问题，尤其是在面

对未见过的药物和蛋白质时。在 DTA 预测任务中，由相似或相同的药物或蛋白质导致的数据冗余可能会导致预测任务变得简单，这可能会混淆方法的性能评估。从实际应用的角度来看，训练集中的大多数药物和蛋白质不会出现在测试集中。模型在遇到未见过的数据时是否仍能表现良好是一个重要挑战。因此，在实验中，我们遵循先前研究（Wang, Wen 等, 2022 年）的设置实施了三种冷启动方案。具体而言，三种不同的分割设置分别如下：

- 药物冷启动：训练集中出现的每种药物都未出现在测试集中。

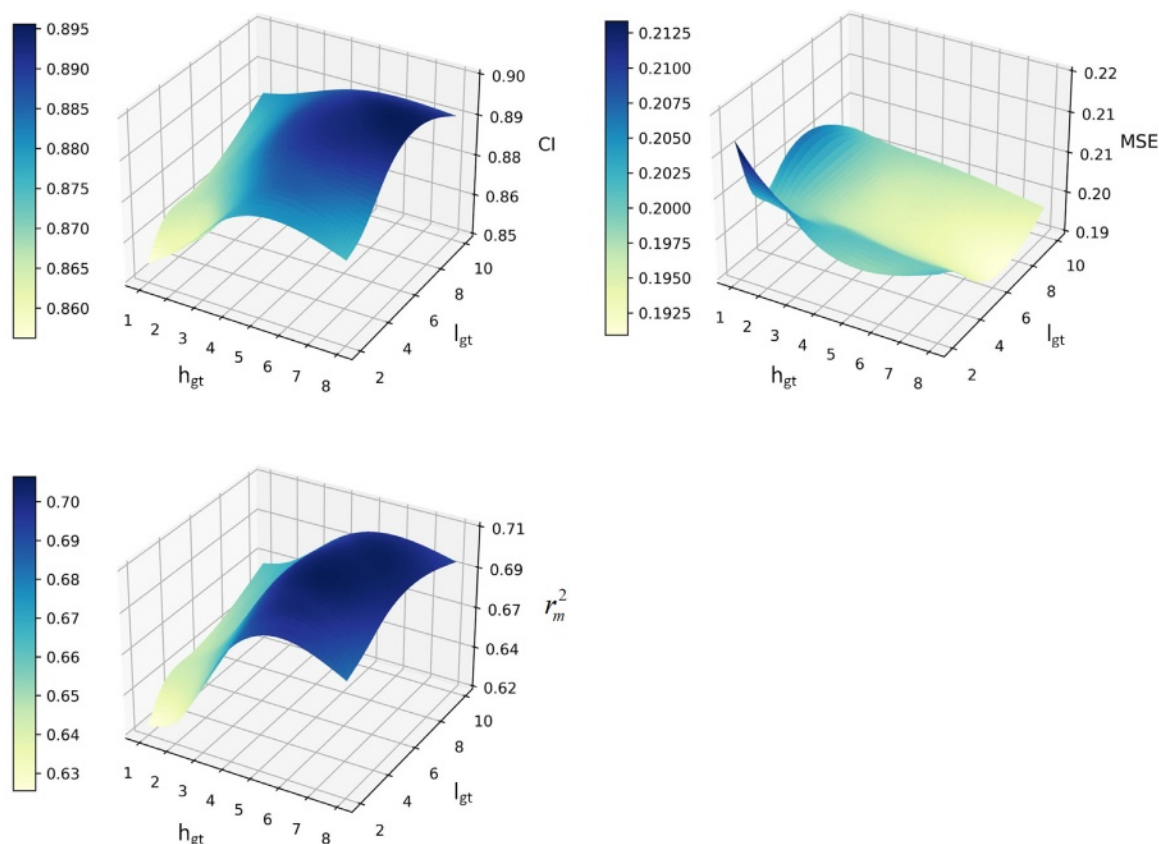


图 5. 在 Davis 数据集上对图变换模型的头和层的参数设置分析。

- 目标冷启动：训练集中出现的每个目标都未出现在测试集中。
- 药物 - 靶点冷启动：同时出现的药物和靶点训练集未出现在测试集中。

这些冷启动方案代表了更现实和复杂的环境，更接近真实世界的应用场景。在戴维斯数据集上的冷启动设置结果如表7和图4所示。我们将 AttentionMGT-DTA 与其他七个基准模型进行了比较。结果表明，所有模型都出现了显著的性能下降，表明这种更现实的情况的复杂性和困难性。我们的模型在三种冷启动设置中获得了最佳的整体性能。在药物冷启动设置中，我们的模型在 CI 和  $R_m^2$  metric 方面显示出优势，分别提高了 3.6% 和 4.9%。在目标冷启动条件下，我们的方法的优点同样明显，CI 和 MSE 值分别提高了 2.5% 和 0.4%。此外，可以看到在药物-目标冷启动方案中，训练期间的药物和目标在测试集中都不存在，与其他基准模型相比，我们的模型显示出令人瞩目的稳定性，在所有三个指标上都优于基准模型。

我们推测，我们的模型在冷启动设置下的稳健性能可以从两个方面来解释：首先，我们基于图的表示方法为药物分子和蛋白质口袋提供了丰富的生物和化学性质特征信息，有助于我们的模型在结合位点附近学习通用和一般的拓扑知识，这可以提高遇到不熟悉数据时的泛化能力。其次，是由于引入了预训练的语言嵌入，提供了通用的蛋白质属性信息，并促进我们的模型通过交互模块学习突出的未见过的蛋白质表示。

这一优点使得 AttentionMGT-DTA 在目标冷启动和药物 - 目标冷启动方案中展现出更出色的优势。总的来说，AttentionMGT-DTA 在冷启动环境中表现出了相对稳定和强大的性能，这证明我们的多模态注意力网络设计在未被发现的 DTA 应用场景中是有效的。

#### 4.5. 参数优化与分析

我们探究了 AttentionMGT-DTA 中超参数的影响：（1）蛋白质图构建的阈值；（2）图变换模块的头数和层数；（3）药物和蛋白质嵌入的维度。我们在戴维斯数据集上进行了参数分析实验。

在我们的蛋白质图构建中，阈值决定了每个残基可以连接的邻接残基数量，这是影响蛋白质残基图复杂性和准确性的重要参数。从经验上讲，增加阈值可以聚合蛋白质的更多结构和拓扑信息。然而，过高的阈值可能导致 GNN 中的连接噪声和过度平滑问题。此外，增加阈值会增加图数据的大小，从而增加模型训练的资源消耗。因此，我们评估了 AttentionMGT-DTA 在不同阈值 {5, 8, 10, 12, 15} 下的性能。表8展示了蛋白质残基图阈值参数设置分析的详细结果。其中，平均边数表示 Davis 数据集中每个蛋白质图的平均边数，平均度表示相应设置下的平均节点度。结果显示，当阈值设置为 10.0 时，我们的模型在所有三个指标上实现了最高性能。之后，我们评估了其影响。

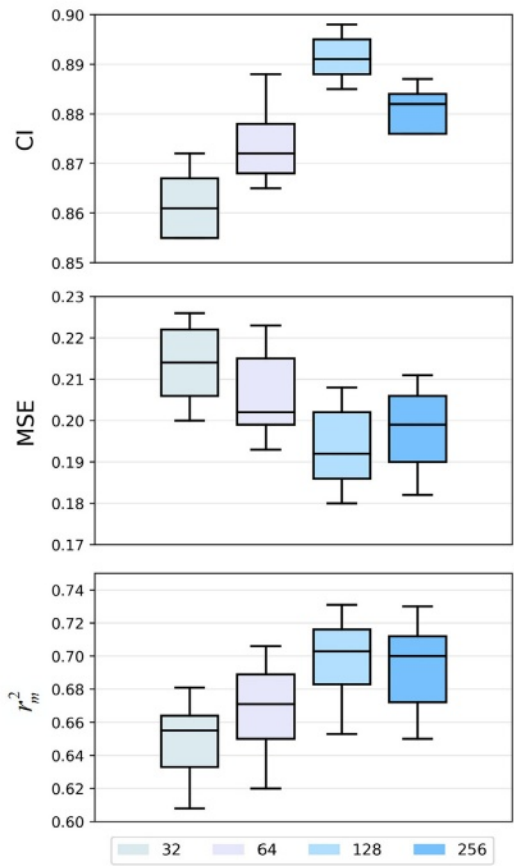


图 6。在戴维斯数据集上对嵌入大小的参数设置分析。

通过网格搜索对图变换器模块中的两个超参数进行优化：注意力头的数量为搜索范围{1, 2, 4, 8}，层数为搜索范围{2, 3, 5, 10}。如图5所示，当注意力头的数量为8，层数为5时，我们的模型的度量值同时达到最大值。最终，考虑到特征嵌入的维度会影响模型学习特征的能力，我们还进行了搜索实验，以探索药物和蛋白质嵌入大小的最佳参数设置，如图6所示，从中我们可以发现不同维度大小之间的评估指标存在显著差异。总体而言，基于上述实验结果和分析，我们已经确定了AttentionMGT-DTA中相应的最优参数设置。

4.6. 消融研究

4.6.1. 模型架构模块的效果

为了验证 AttentionMGT-DTA 中每个模块的贡献和有效性，我们在 Davis 数据集上进行了消融研究。我们通过移除预训练的蛋白质嵌入、交叉注意力和联合注意力模块进行了消融实验。具有不同模块的模型的性能列于表 9 中。具体而言，第一个模型AttentionMGT-DTA<sub>CONCAT</sub>concatenated使用蛋白质图嵌入和预训练序列嵌入，而不是使用交叉注意力机制。AttentionMGT-DTA<sub>SINGLE</sub>是没有预训练蛋白质嵌入特征的模型。对于第三个模型AttentionMGT-DTA<sub>MAX</sub>，通过在图变换器后添加最大化池化层来移除联合注意力机制，并且预训练嵌入也被聚合为一个相同的向量。

表9

在戴维斯数据集上的消融研究结果。

模型	R2M	持续集成	均方误差
AttentionMGT-DTA <sub>CONCAT</sub>	0.663 (0.026)	0.876 (0.006)	0.206 (0.006)
AttentionMGT-DTA <sub>SINGLE</sub>	0.678 (0.021)	0.889 (0.005)	0.198 (0.006)
AttentionMGT-DTA <sub>MAX</sub>	0.667 (0.021)	0.886 (0.007)	0.198 (0.006)
AttentionMGT-DTA <sub>MEAN</sub>	0.673 (0.030)	0.886 (0.009)	0.199 (0.007)
AttentionMGT-DTA <sub>PRE</sub>	0.675 (0.025)	0.887 (0.006)	0.196 (0.012)
注意：MGT - DTA	0.699 (0.027)	0.891 (0.005)	0.193 (0.010)

通过最大化函数进行维度分析。类似地，AttentionMGT-DTA<sub>MEAN</sub> was 该模型使用均值池化方法来替代联合注意力模块。

如表9所示，AttentionMGT-DTA与变体模型相比表现最佳。去除交叉注意力模块显著降低了模型的预测性能，这清楚地表明了多模态交互的重要性。对于预训练嵌入的消融研究，结果表明，我们的模型引入的1D序列嵌入提供了高效的高级蛋白质序列信息，从而提高了性能。此外，AttentionMGT-DTA还优于第三个和第四个变体模型，表明联合注意力机制增强了我们模型的性能。我们推测这主要是因为我们的模型在药物和目标数据之间的信息交互能力，优于传统的池化和连接策略，有利于DTA预测。总体而言，消融实验表明，我们的模型中的模态和数据交互模块对于提高预测性能是有效的。

4.6.2. AlphaFold 结构的效果

如上所述，蛋白质晶体结构的可用性也是影响最终性能的一个主要因素。考虑到PDB数据库中部分蛋白质结构信息缺失或不可用，我们因此选择了来自AlphaFold数据库的蛋白质结构，而不是PDB数据库。我们进一步研究了AlphaFold结构对整体性能的影响，通过将AlphaFold结构替换为可用的PDB结构。具体来说，在Davis数据集中，可用的PDB蛋白质结构数量为275，而由AlphaFold补充的缺失结构数量为86。如表9所示，模型在CI、MSE和R<sup>2</sup><sub>M</sub>index方面的性能分别提高了0.4%、0.3%和2.4%。这种改进可能是由于数据集中存在更多结构相似的蛋白质，使得我们的模型能够学习药物-靶点对之间的复杂结合特性和特征，而两种数据库结构的混合存在可能会影响模型的学习能力。为此，我们证明了利用高品质的 AlphaFold 预测结构增强了所提出模型的性能，并使我们的方法优于现有的基准。

4.7. 解读与可视化

AttentionMGT-DTA 中的联合注意力模块可用于分析哪些蛋白质结合位点更有可能与药物分子结合。通过输入药物图和蛋白质图的表示，我们的模型可以生成注意力矩阵，该矩阵展示了在药物 - 靶点结合中每个蛋白质残基和药物原子的重要性。注意力矩阵可以为这些概率提供生物学上合理的解释，这是我们模型的一个优势。图 7 展示了所提出模型的注意力可视化的示例。为了说明该模型的可解释性，我们从 PDB (Burley 等人, 2018 年) 中选择了两个复合物进行结合可视化分析。我们对蛋白质残基和药物原子的高权重位置进行了着色。特别是，着色的残基突出了具有高注意力分数的蛋白质结合口袋的位置。



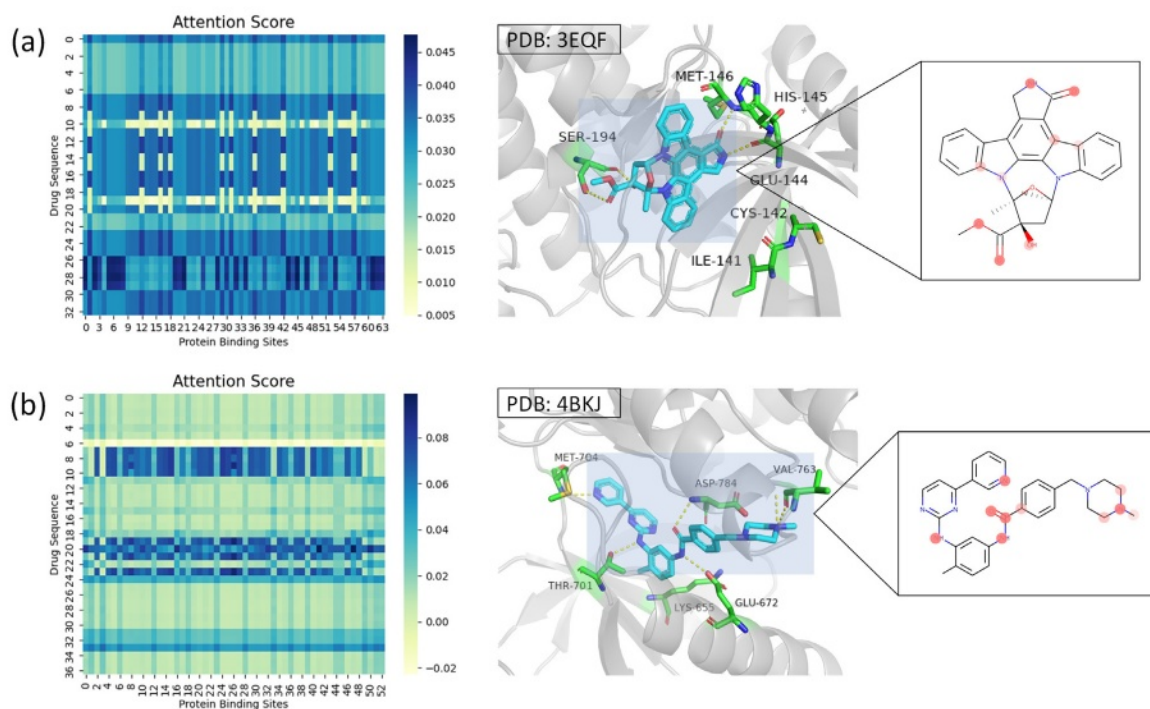


图 7. 药物-靶点相互作用的注意力可视化。左：注意力矩阵的热图。中：药物和突出显示的残基分别用蓝色和绿色表示。右：带有突出显示原子的药物结构，用红色表示。(a) 3EQF 的注意力可视化。(b) 4BKJ 的注意力可视化。

并且红色突出显示了重点药物原子。注意力得分分别通过对蛋白质维度和药物维度的注意力矩阵取平均值来获得。

对于图7中的蛋白质MAP2K1 (Uniport ID: Q02750)，模型识别的关键残基包括ILE141、CYS142、GLU144、HIS145、MET146和SER194，这些残基部分重叠于复合物结构 (PDB ID: 3EQF) 中观察到的结合位点残基。类似地，观察到具有高注意力值的药物原子位于蛋白质口袋内或周围。对于图7中的蛋白质DDR1 (UniProt ID: Q08345)，关键残基 (LYS655、GLU672、THR701、MET704、VAL763、ASP784) 和分子原子也与共晶体复合物 (PDB ID: 4BKJ) 中观察到的结合相似。总体而言，我们的模型捕获的大多数残基位于结合位点，但仍然存在一些错误的预测。结果表明，注意力机制可以通过学习重要的蛋白质残基和药物原子来提取有意义的结合信息。此外，结果表明，所提出的模型可以帮助我们准确和全面地理解DTA，这有利于研究人员研究目标蛋白质与药物之间的结合和相互作用机制。

## 5. 结论

在本文中，我们提出了一种用于 DTA 预测的新颖模型，名为 AttentionMGT-DTA，该模型基于注意力机制来捕获各种独立模态之间的关系。AttentionMGT-DTA 通过基于图和注意力的编码器采用多模态蛋白质残基级特征和药物原子级特征。同时，我们的模型可以使用药物和蛋白质之间的信息融合模块来学习注意力矩阵，这能够为生物学意义提供重要的解释。在公共数据集上的实验结果表明，我们的模型比现有模型表现更好。当遇到未知的药物和蛋白质时，AttentionMGT-DTA 也是强大且有效的。此外，我们模型的优势也表明，从 AlphaFold 数据库预测的蛋白质结构具有高质量，能够为下游任务提供准确和有效的结构信息。

尽管我们的模型已被证明具有出色的性能，但仍有很大的改进空间。我们只使用了一种模态，即二维分子图来表示药物，而没有考虑其他表示形式，如序列和指纹。在未来的工作中，我们将专注于在多模态深度学习模型中集成和融合药物和靶点。此外，我们工作中用于确定蛋白质结合口袋的算法有时不准确，导致数据集中蛋白质的口袋不匹配，这在一定程度上会影响性能。因此，引入并应用更多的方法来寻找蛋白质的隐蔽口袋，使蛋白质中存在更多可以与药物结合的位点，是另一个未来的方向，以提高性能并增强泛化能力。

## 代码可用性

该代码在 GitHub 上免费提供：<https://github.com/JK-Liu7/AttentionMGT-DTA>。

## 利益冲突声明

作者声明不存在利益冲突。

## 数据可用性

论文中给出了 Github 链接。

## 致谢

本研究得到了中国国家自然科学基金 (62073231、62176175、62172076)、国家研究项目 (2020YFC2006602)、江苏省计算机信息处理技术重点实验室 (苏州大学 KJS2166)、江苏省大数据智能工程实验室开放课题基金 (SDGC2157)、江苏省研究生研究与实践创新计划、浙江省自然科学基金 (编号: LY23F020003) 以及中国衢州市政府 (编号: 2023D038) 的支持。

## 参考文献

- Bagherian, M., Sabeti, E., Wang, K., Sartor, M. A., Nikolovska-Coleska, Z., & Najarian, K. (2020). Machine learning approaches and databases for prediction of drug–target interaction: A survey paper. *Briefings in Bioinformatics*, 22(1), 247–269. <http://dx.doi.org/10.1093/bib/bbz157>.
- Bahi, M., & Batouche, M. (2021). Convolutional neural network with stacked autoencoders for predicting drug–target interaction and binding affinity. *International Journal of Data Mining, Modelling and Management*, 13(1–2), 81–113. <http://dx.doi.org/10.1504/IJDDMM.2021.112914>.
- Bepler, T., & Berger, B. (2021). Learning the protein language: Evolution, structure, and function. *Cell Systems*, 12(6), 654–669. <http://dx.doi.org/10.1016/j.cels.2021.05.017>.
- Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., et al. (2018). RCSB Protein Data Bank: Biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research*, 47(D1), D464–D474. <http://dx.doi.org/10.1093/nar/gky1004>.
- Chao, W., & Quan, Z. (2021). A machine learning method for differentiating and predicting human-infective coronavirus based on physicochemical features and composition of the spike protein. *Chinese Journal of Electronics*, 30(5), 815–823. <http://dx.doi.org/10.1049/cje.2021.06.003>.
- Chen, L., Tan, X., Wang, D., Zhong, F., Liu, X., Yang, T., et al. (2020). TransformerCPI: Improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, 36(16), 4406–4414. <http://dx.doi.org/10.1093/bioinformatics/btaa524>.
- Cloninger, A., & Klock, T. (2021). A deep network construction that adapts to intrinsic dimensionality beyond the domain. *Neural Networks*, 141, 404–419. <http://dx.doi.org/10.1016/j.neunet.2021.06.004>.
- Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., et al. (2011). Comprehensive analysis of kinase inhibitor selectivity. *Nature biotechnology*, 29(11), 1046–1051. <http://dx.doi.org/10.1038/nbt.1990>.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. [arXiv:1810.04805](https://arxiv.org/abs/1810.04805).
- Dhakal, A., McKay, C., Tanner, J. J., & Cheng, J. (2021). Artificial intelligence in the prediction of protein–ligand interactions: Recent advances and future directions. *Briefings in Bioinformatics*, 23(1), bbab476. <http://dx.doi.org/10.1093/bib/bbab476>.
- Ding, Y., Guo, F., Tiwari, P., & Zou, Q. (2023). Identification of drug-side effect association via multi-view semi-supervised sparse model. *IEEE Transactions on Artificial Intelligence*, 1–12. <http://dx.doi.org/10.1109/TAI.2023.3314405>.
- Ding, Y., Tang, J., & Guo, F. (2020a). Identification of drug–target interactions via dual Laplacian regularized least squares with multiple kernel fusion. *Knowledge-Based Systems*, 204, Article 106254. <http://dx.doi.org/10.1016/j.knsys.2020.106254>.
- Ding, Y., Tang, J., & Guo, F. (2020b). Identification of drug–target interactions via fuzzy bipartite local model. *Neural Computing and Applications*, 32, 10303–10319. <http://dx.doi.org/10.1007/s00521-019-04569-z>.
- Ding, Y., Tang, J., & Guo, F. (2021). Identification of drug–target interactions via multi-view graph regularized link propagation model. *Neurocomputing*, 461, 618–631. <http://dx.doi.org/10.1016/j.neucom.2021.05.100>.
- Ding, Y., Tang, J., Guo, F., & Zou, Q. (2022). Identification of drug–target interactions via multiple kernel-based triple collaborative matrix factorization. *Briefings in Bioinformatics*, 23(2), <http://dx.doi.org/10.1093/bib/bbab582>, bbab582.
- Ding, Y., Tiwari, P., Guo, F., & Zou, Q. (2022). Shared subspace-based radial basis function neural network for identifying ncRNAs subcellular localization. *Neural Networks*, 156, 170–178. <http://dx.doi.org/10.1016/j.neunet.2022.09.026>.
- Dwivedi, V. P., & Bresson, X. (2020). A generalization of transformer networks to graphs. [arXiv:2012.09699](https://arxiv.org/abs/2012.09699).
- Dwivedi, V. P., Joshi, C. K., Luu, A. T., Laurent, T., Bengio, Y., & Bresson, X. (2020). Benchmarking graph neural networks. [http://dx.doi.org/10.48550/arXiv.2003.00982](https://arxiv.org/abs/10.48550/arXiv.2003.00982), [arXiv preprint arXiv:2003.00982](https://arxiv.org/abs/2003.00982).
- ElAbd, H., Bromberg, Y., Hoarfrost, A., Lenz, T., Franke, A., & Wendorff, M. (2020). Amino acid encoding for deep learning applications. *BMC Bioinformatics*, 21, 1–14. <http://dx.doi.org/10.1186/s12859-020-03546-x>.
- Ezzat, A., Wu, M., Li, X.-L., & Kwok, C.-K. (2018). Computational prediction of drug–target interactions using chemogenomic approaches: An empirical survey. *Briefings in Bioinformatics*, 20(4), 1337–1357. <http://dx.doi.org/10.1093/bib/bby002>.
- Feng, Q., Dueva, E., Cherkasov, A., & Ester, M. (2019). PADME: A deep learning-based framework for drug–target interaction prediction. [arXiv:1807.09741](https://arxiv.org/abs/1807.09741).
- Gönen, M., & Heller, G. (2005). Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92(4), 965–970.
- Huang, L., Lin, J., Liu, R., Zheng, Z., Meng, L., Chen, X., et al. (2022). CoaDTI: Multi-modal co-attention based framework for drug–target interaction annotation. *Briefings in Bioinformatics*, 23(6), bbac446. <http://dx.doi.org/10.1093/bib/bbac446>.
- Huang, K., Xiao, C., Glass, L. M., & Sun, J. (2020). MolTrans: Molecular interaction transformer for drug–target interaction prediction. *Bioinformatics*, 37(6), 830–836. <http://dx.doi.org/10.1093/bioinformatics/btaa880>.
- Jiang, D., Hsieh, C.-Y., Wu, Z., Kang, Y., Wang, J., Wang, E., et al. (2021). Interac-tionGraphNet: A novel and efficient deep graph representation learning framework for accurate protein–ligand interaction predictions. *Journal of Medicinal Chemistry*, 64(24), 18209–18232. <http://dx.doi.org/10.1021/acs.jmedchem.1c01830>, PMID: 34878785.
- Jiang, M., Li, Z., Zhang, S., Wang, S., Wang, X., Yuan, Q., et al. (2020). Drug–target affinity prediction using graph neural network and contact maps. *RSC Advances*, 10, 20701–20712. <http://dx.doi.org/10.1039/D0RA02297G>.
- Jumper, J., Evans, R., Pritzel, A., Green, T., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <http://dx.doi.org/10.1038/s41586-021-03819-2>.
- Karimi, M., Wu, D., Wang, Z., & Shen, Y. (2019). DeepAffinity: Interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics*, 35(18), 3329–3338. <http://dx.doi.org/10.1093/bioinformatics/btz111>.
- Karimi, M., Wu, D., Wang, Z., & Shen, Y. (2021). Explainable deep relational networks for predicting compound–protein affinities and contacts. *Journal of Chemical Information and Modeling*, 61(1), 46–66. <http://dx.doi.org/10.1021/acs.jcim.0c00866>, PMID: 33347301.
- Kimber, T. B., Chen, Y., & Volkamer, A. (2021). Deep learning in virtual screening: Recent applications and developments. *International Journal of Molecular Sciences*, 22(9), <http://dx.doi.org/10.3390/ijms22094435>.
- Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. [arXiv: 1412.6980](https://arxiv.org/abs/1412.6980).
- Lee, I., Keum, J., & Nam, H. (2019). DeepConv-DTI: Prediction of drug–target interactions via deep learning with convolution on protein sequences. *PLoS Computational Biology*, 15(6), 1–21. <http://dx.doi.org/10.1371/journal.pcbi.1007129>.
- Li, S., Wan, F., Shu, H., Jiang, T., Zhao, D., & Zeng, J. (2020). MONN: A multi-objective neural network for predicting compound–protein interactions and affinities. *Cell Systems*, 10(4), 308–322.e11. <http://dx.doi.org/10.1016/j.cels.2020.03.002>.
- Li, F., Zhang, Z., Guan, J., & Zhou, S. (2022a). Effective drug–target interaction pre-diction with mutual interaction neural network. *Bioinformatics*, 38(14), 3582–3589. <http://dx.doi.org/10.1093/bioinformatics/btac377>.
- Li, F., Zhang, Z., Guan, J., & Zhou, S. (2022b). Effective drug–target interaction pre-diction with mutual interaction neural network. *Bioinformatics*, 38(14), 3582–3589. <http://dx.doi.org/10.1093/bioinformatics/btac377>.
- Li, T., Zhao, X.-M., & Li, L. (2022). Co-VAE: Drug-target binding affinity prediction by co-regularized variational autoencoders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12), 8861–8873. <http://dx.doi.org/10.1109/TPAMI.2021.3120428>.
- Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637), 1123–1130. <http://dx.doi.org/10.1126/science.ade2574>.
- Mao, T., Shi, Z., & Zhou, D. (2021). Theory of deep convolutional neural networks III: Approximating radial functions. *Neural Networks*, 144, 778–790. <http://dx.doi.org/10.1016/j.neunet.2021.09.027>.
- Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10), 2319–2327. <http://dx.doi.org/10.1002/jcc.21787>.
- Newman, D. J., & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, 83(3), 770–803. <http://dx.doi.org/10.1021/acs.jnatprod.9b01285>, PMID: 32162523.
- Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2020). GraphDTA: Predicting drug–target binding affinity with graph neural networks. *Bioinformatics*, 37(8), 1140–1147. <http://dx.doi.org/10.1093/bioinformatics/btaa921>.
- Nguyen, T. M., Nguyen, T., Le, T. M., & Tran, T. (2022a). GEFA: Early fusion approach in drug–target affinity prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2), 718–728. <http://dx.doi.org/10.1109/TCBB.2021.3094217>.
- Nguyen, T. M., Nguyen, T., Le, T. M., & Tran, T. (2022b). GEFA: Early fusion approach in drug–target affinity prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(2), 718–728. <http://dx.doi.org/10.1109/TCBB.2021.3094217>.
- Öztürk, H., Özgür, A., & Ozkirimli, E. (2018). DeepDTA: Deep drug–target binding affinity prediction. *Bioinformatics*, 34(17), i821–i829. <http://dx.doi.org/10.1093/bioinformatics/bty593>.
- Öztürk, H., Ozkirimli, E., & Özgür, A. (2019). WideDTA: Prediction of drug–target binding affinity. [arXiv:1902.04166](https://arxiv.org/abs/1902.04166).
- Pandey, M., Radaeva, M., Mslati, H., Garland, O., Fernandez, M., Ester, M., et al. (2022). Ligand binding prediction using protein structure graphs and residual graph attention networks. *Molecules*, 27(16), <http://dx.doi.org/10.3390/molecules27165114>, URL <https://www.mdpi.com/1420-3049/27/16/5114>.
- Qian, Y., Ding, Y., Zou, Q., & Guo, F. (2022). Identification of drug-side effect association via restricted Boltzmann machines with penalized term. *Briefings in Bioinformatics*, 23(6), <http://dx.doi.org/10.1093/bib/bbac458>, bbac458.
- Rifaioğlu, A. S., Nalbat, E., Atalay, V., Martin, M. J., Cetin-Atalay, R., & Doğan, T. (2020). DEEPScreen: High performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chemical Science*, 11, 2531–2557. <http://dx.doi.org/10.1039/C9SC03414E>.
- Saberi Fathi, S. M., & Tuszyński, J. A. (2014). A simple method for finding a protein’s ligand-binding pockets. *BMC Structural Biology*, 14(1), 1–9. <http://dx.doi.org/10.1186/1472-6807-14-18>.

- Shen, C., Zhang, X., Deng, Y., Gao, J., Wang, D., Xu, L., et al. (2022). Boosting protein–ligand binding pose prediction and virtual screening based on residue–atom distance likelihood potential and graph transformer. *Journal of Medicinal Chemistry*, 65(15), 10691–10706. <http://dx.doi.org/10.1021/acs.jmedchem.2c00991>, PMID: 35917397.
- Tang, J., Szwajda, A., Shakyawar, S., Xu, T., Hintsanen, P., Wennerberg, K., et al. (2014). Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. *Journal of Chemical Information and Modeling*, 54(3), 735–743. <http://dx.doi.org/10.1021/ci400709d>, PMID: 24521231.
- Torng, W., & Altman, R. B. (2019). Graph convolutional neural networks for predicting drug–target interactions. *Journal of Chemical Information and Modeling*, 59(10), 4131–4149. <http://dx.doi.org/10.1021/acs.jcim.9b00628>, PMID: 31580672.
- Tsubaki, M., Tomii, K., & Sese, J. (2018). Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, 35(2), 309–318. <http://dx.doi.org/10.1093/bioinformatics/bty535>.
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2021). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein–sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1), D439–D444. <http://dx.doi.org/10.1093/nar/gkab1061>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. [arXiv:1706.03762](https://arxiv.org/abs/1706.03762).
- Wang, J., Hu, J., Sun, H., Xu, M., Yu, Y., Liu, Y., et al. (2022). MGPLI: Exploring multi-granular representations for protein–ligand interaction prediction. *Bioinformatics*, 38(21), 4859–4867. <http://dx.doi.org/10.1093/bioinformatics/btac597>.
- Wang, L., Liu, H., Liu, Y., Kurtin, J., & Ji, S. (2023). Learning hierarchical protein representations via complete 3D graph networks. In *The eleventh international conference on learning representations*. URL <https://openreview.net/forum?id=9X-hgLDLYkQ>.
- Wang, H., Tang, J., Ding, Y., & Guo, F. (2021). Exploring associations of non-coding RNAs in human diseases via three-matrix factorization with hypergraph-regular terms on center kernel alignment. *Briefings in Bioinformatics*, 22(5), <http://dx.doi.org/10.1093/bib/bbaa409>, bbaa409.
- Wang, J., Wen, N., Wang, C., Zhao, L., & Cheng, L. (2022). ELECTRA-DTA: A new compound–protein binding affinity prediction model based on the contextualized sequence encoding. *Journal of cheminformatics*, 14(1), 1–14. <http://dx.doi.org/10.1186/s13321-022-00591-x>.
- Wang, Y., Zhai, Y., Ding, Y., & Zou, Q. (2023). SBSM-Pro: Support bio-sequence machine for proteins. <http://dx.doi.org/10.48550/arXiv.2308.10275>, [arXiv preprint arXiv:2308.10275](https://arxiv.org/abs/2308.10275).
- Wang, Z., Zhang, Q., HU, S.-W., Yu, H., Jin, X., Gong, Z., et al. (2023). Multi-level protein structure pre-training via prompt learning. In *The eleventh international conference on learning representations*. URL <https://openreview.net/forum?id=XGagtiJ8XC>.
- Wang, P., Zheng, S., Jiang, Y., Li, C., Liu, J., Wen, C., et al. (2022). Structure-aware multimodal deep learning for drug–protein interaction prediction. *Journal of Chemical Information and Modeling*, 62(5), 1308–1317. <http://dx.doi.org/10.1021/acs.jcim.2c00060>, PMID: 35200015.
- Wang, K., Zhou, R., Li, Y., & Li, M. (2021). DeepDTAF: A deep learning method to predict protein–ligand binding affinity. *Briefings in Bioinformatics*, 22(5), bbab072. <http://dx.doi.org/10.1093/bib/bbab072>.
- Wouters, O. J., McKee, M., & Luyten, J. (2020). Research and development costs of new drugs—Reply. *JAMA*, 324(5), 518. <http://dx.doi.org/10.1001/jama.2020.8651>.
- Wu, Y., Gao, M., Zeng, M., Zhang, J., & Li, M. (2022). BridgeDPI: A novel Graph Neural Network for predicting drug–protein interactions. *Bioinformatics*, 38(9), 2571–2578. <http://dx.doi.org/10.1093/bioinformatics/btac155>.
- Wu, Z., Jiang, D., Wang, J., Hsieh, C.-Y., Cao, D., & Hou, T. (2021). Mining toxicity information from large amounts of toxicity data. *Journal of Medicinal Chemistry*, 64(10), 6924–6936. <http://dx.doi.org/10.1021/acs.jmedchem.1c00421>, PMID: 33961429.
- Wu, H., Ling, H., Gao, L., Fu, Q., Lu, W., Ding, Y., et al. (2021). Empirical potential energy function toward ab initio folding g protein-coupled receptors. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(5), 1752–1762. <http://dx.doi.org/10.1109/TCBB.2020.3008014>.
- Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? [arXiv:1810.00826](https://arxiv.org/abs/1810.00826).
- Yang, H., Ding, Y., Tang, J., & Guo, F. (2021). Drug–disease associations prediction via multiple kernel-based dual graph regularized least squares. *Applied Soft Computing*, 112, Article 107811. <http://dx.doi.org/10.1016/j.asoc.2021.107811>.
- Yang, Z., Zhong, W., Zhao, L., & Chen, C. Y.-C. (2021). ML-DTI: Mutual learning mechanism for interpretable drug–target interaction prediction. *The Journal of Physical Chemistry Letters*, 12(17), 4247–4261. <http://dx.doi.org/10.1021/acs.jpclett.1c00867>, PMID: 33904745.
- Yang, Z., Zhong, W., Zhao, L., & Yu-Chian Chen, C. (2022). MGraphDTA: Deep multi-scale graph neural network for explainable drug–target binding affinity prediction. *Chemical Science*, 13, 816–833. <http://dx.doi.org/10.1039/D1SC05180F>.
- Yazdani-Jahromi, M., Yousefi, N., Tayebi, A., Kolanthai, E., Neal, C. J., Seal, S., et al. (2022). AttentionSiteDTI: An interpretable graph-based model for drug–target interaction prediction using NLP sentence-level relation classification. *Briefings in Bioinformatics*, 23(4), bbac272. <http://dx.doi.org/10.1093/bib/bbac272>.
- Zhang, Z., Chen, L., Zhong, F., Wang, D., Jiang, J., Zhang, S., et al. (2022). Graph neural network approaches for drug–target interactions. *Current Opinion in Structural Biology*, 73, Article 102327. <http://dx.doi.org/10.1016/j.sbi.2021.102327>.
- Zhang, F., Song, H., Zeng, M., Wu, F.-X., Li, Y., Pan, Y., et al. (2021). A deep learning framework for gene ontology annotations with sequence- and network-based information. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(6), 2208–2217. <http://dx.doi.org/10.1109/TCBB.2020.2968882>.
- Zhang, Y., Tiwari, P., Song, D., Mao, X., Wang, P., Li, X., et al. (2021). Learning interaction dynamics with an interactive LSTM for conversational sentiment analysis. *Neural Networks*, 133, 40–56.
- Zhang, Z., Xu, M., Jamasb, A., Chenthamarakshan, V., Lozano, A., Das, P., et al. (2023). Protein representation learning by geometric structure pretraining. [arXiv: 2203.06125](https://arxiv.org/abs/2203.06125).
- Zhao, Q., Xiao, F., Yang, M., Li, Y., & Wang, J. (2019). AttentionDTA: Prediction of drug–target binding affinity using attention model. In *2019 IEEE international conference on bioinformatics and biomedicine* (pp. 64–69). <http://dx.doi.org/10.1109/BIBM47256.2019.8983125>.
- Zhao, Q., Zhao, H., Zheng, K., & Wang, J. (2021). HyperAttentionDTI: Improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*, 38(3), 655–662. <http://dx.doi.org/10.1093/bioinformatics/btab715>.
- Zheng, S., Li, Y., Chen, S., Xu, J., & Yang, Y. (2020). Predicting drug–protein interaction using quasi-visual question answering system. *Nature Machine Intelligence*, 2(2), 134–140. <http://dx.doi.org/10.1038/s42256-020-0152-y>.