

## · 综述 ·

## 人工智能算法用于药物研发的研究进展

杨双萌<sup>1</sup>, 于江<sup>2</sup>, 侯文彬<sup>2</sup>, 赵倩<sup>3\*</sup>, 李祎亮<sup>2\*</sup>

1. 天津中医药大学, 天津 301617

2. 中国医学科学院 北京协和医学院 放射医学研究所, 天津 300192

3. 天津市南开区王顶堤医院, 天津 300190

**摘要:** 人工智能算法包含机器学习算法和深度学习算法, 可应用于药物靶标发现、先导化合物的发现与优化、候选药物的确定、成药性优化。人工智能算法通过丰富的大数据系统学习可以实现模型的建立和高通量虚拟计算, 应用于药物研发中能够在一定程度上缩短研发周期、降低投入成本, 进而提高研发成功率。对机器学习算法、深度学习算法应用于药物研发中的研究进展进行阐述, 以期为人工智能技术与药物研发相结合的进一步发展提供参考。

**关键词:** 药物研发; 人工智能算法; 机器学习算法; 深度学习算法; 药物靶标; 先导化合物; 候选药物; 成药性

**中图分类号:** R95      **文献标志码:** A      **文章编号:** 1674-5515(2023)12-3150-11

**DOI:** 10.7501/j.issn.1674-5515.2023.12.042

## Research progress on artificial intelligence algorithms for drug discovery

YANG Shuang-meng<sup>1</sup>, YU Jiang<sup>2</sup>, HOU Wen-bin<sup>2</sup>, ZHAO Qian<sup>3</sup>, LI Yi-liang<sup>2</sup>

1. Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China

2. Institute of Radiation Medicine Chinese Academy of Medical Sciences &amp; Peking Union Medical College, Tianjin 300192, China

3. Tianjin Nankai District Wangdingdi Hospital, Tianjin 300190, China

**Abstract:** Artificial intelligence algorithms include machine learning algorithms and deep learning algorithms, which can be applied to discover drug targets, discover and optimize lead compounds, determine candidate drugs, and optimize druggability. Artificial intelligence algorithms can achieve model building and high-throughput virtual computing through complex big data system learning. When applied in drug research and development, it can shorten the research and development cycle to a certain extent, reduce input costs, and thereby improve the success rate of research and development. In this article, the research progress of machine learning algorithms and deep learning algorithms applied in drug development was reviewed, in order to provide reference for the further development of the combination of artificial intelligence technology and drug development.

**Key words:** drug discovery; artificial Intelligence algorithms; machine learning algorithms; deep learning algorithms; drug target; lead compound; candidate drug; druggability

人工智能是计算机科学的一个分支, 指计算机从已有的数据集中进行学习的能力, 能够促进大量结构化和非结构化数据的处理和分析<sup>[1-4]</sup>。人工智能算法包含机器学习算法和深度学习算法<sup>[5-8]</sup>。机器学习是指在没有显式编程的情况下逼近给定训练数据的输入与输出之间的映射函数, 从而能够对新数

据进行预测, 而深度学习是一种训练神经网络的具体方法, 就像大脑中神经元对特定的视觉输入做出反应一样。机器学习是使用传统统计算法学习映射函数, 深度学习是使用神经网络结构学习映射<sup>[9]</sup>。药物研发的过程漫长而复杂, 并且会受到多种因素影响<sup>[10]</sup>, 通常包括药物靶标发现、先导化合物的发

收稿日期: 2023-09-21

基金项目: 国家自然科学基金项目资助 (82104012, 82202950, 82303681)

作者简介: 杨双萌, 女, 中药学专业硕士生。E-mail: shuangmengyang@163.com

\*通信作者: 赵倩 E-mail: zhaoqian\_nk@126.com

李祎亮 E-mail: liyiliang@irm-cams.ac.cn

现与优化、候选药物的确定、成药性优化等方面<sup>[11]</sup>。目前无论是机器学习还是深度学习,都能够通过丰富的大数据系统进行学习,已经在药物研发中得到广泛应用<sup>[10, 12-14]</sup>。人工智能算法通过丰富的大数据系统学习可以实现模型的建立和高通量虚拟计算,应用于药物研发中能够在一定程度上缩短研发周期、降低投入成本,进而提高研发成功率。本文对机器学习算法、深度学习算法应用于药物研发中的研究进展进行阐述,以期为人工智能技术与药物研发相结合的进一步发展提供参考。

## 1 机器学习在药物研发中的应用

机器学习算法通过数学建模来找到特定活动或化合物的分类及其特征之间的关联。目前已经得到可用的机器学习方法有朴素贝叶斯、支持向量机、递归分区、k 最邻近、决策树、随机森林等<sup>[15-18]</sup>。

### 1.1 机器学习应用于药物靶标发现

药物靶标是指体内具有药效功能并能被药物作用的生物大分子,即药物发挥疗效的靶点,如某些蛋白质和核酸等生物大分子。一个好的药物靶点需要与疾病表型相关,并且适合用于治疗调节<sup>[19]</sup>。确定靶标分子是现代药物开发的基础。识别药物与靶标的相互作用将极大地缩小药物搜索范围,因此是药物发现的关键第一步<sup>[20-21]</sup>。据报道,目前已有 30 多个药物靶标发现平台,能够为研究人员提供丰富的数据库、图形界面,便于预测药物在靶点的治疗效果和潜在的不良反应<sup>[22]</sup>。

**1.1.1 基于靶标(蛋白质可药性)** 鉴定蛋白质靶标活性是药物发现的步骤之一,确定蛋白质结合药物以调节其功能的能力称为可药性。由于可药性衡量需要大量时间和资源,因而导致大量药物发现的失败<sup>[22]</sup>。因此人工智能技术对于基于靶点的药物靶标发现具有重要意义。

Kana 等<sup>[23]</sup>提出基于“eFindSite”的药效预测算法,通过使用有监督的机器学习对蛋白质的可药性进行预测,从而实现了人类蛋白质组药物分析的口袋可药性预测。Cavasotto 等<sup>[24]</sup>通过实验结构数据或同源模型对 SARS-CoV-2 蛋白质组进行了全面电子药物敏感性评估,包括其所有非结构蛋白、结构蛋白和辅助蛋白都进行了深入的功能、结构和可药性评估,确定了整个蛋白质组中潜在的可药物变构和蛋白质-蛋白质相互作用(PPI)位点,为药物发现提供丰富的靶点。Lee 等<sup>[25]</sup>开发的“PharmMaker”软件工具,从可药性模拟、药效团

建模、化合物库虚拟筛选等进行了探索,通过使用 Prody 中的 DruGUI 模块从可药性模拟中识别热点、药物作用部位高亲和力残基的筛选、基于发生频率的残基-探针相互作用排序成功构建药效团模型,且此软件可从网页中获取应用于实验。

**1.1.2 基于配体(小分子化合物)** 现代计算机数据库中能够免费获得的分子数量庞大,因此计算机辅助配体发现成为早期配体发现的替代方案,也被称为虚拟筛选。基于配体的虚拟筛选侧重于没有靶标结构情况下的配体结构及其理化性质,广义上可成为基于配体的相似性搜索或属性预测的技术<sup>[26]</sup>。目前已有能够对分子特征进行评估的机器学习框架和算法,可根据特定评估指标或累积得分判断分子匹配度,对活性化合物开发起到参考作用。

Chaube 等<sup>[27]</sup>开发了用于预测镧系元素与不同分子配体结合亲和力的机器学习框架,通过 6 种机器学习算法对已有结合亲和力数据集进行训练,并进行交叉验证,随后进行全面的特征工程和特征重要性分析,以确定相关性最佳的分子、金属和溶剂特征,同时给出特征空间的维度评估性能指标。Da'adoosh 等<sup>[28]</sup>提出应用机器学习算法构建基于配体的多翼排序模型,包含 4 个物理化学性质反位积和的过滤器组成的模型,能够预测具有高度多样化结构的新型生物活性物质,并结合对接和几何过滤对高活性分子进行筛选。分子通过过滤器产生的分数最终累积为 MBI 分数,从而发现具有新支架的新配体,且具有更优的、能够避免产生与已知结构相似化合物的能力。Sun 等<sup>[29]</sup>使用抑制剂模型和分子对接的方法进行数据库筛选,从而获得潜在的抑制剂亲和力数据。此研究证明了归纳逻辑编程方法在基于配体的化合物虚拟筛选中的应用,该方法基于实验验证的化合物分子结构和所需靶标的知识背景进行机器学习,并且归纳逻辑编程方法,可以由研究中的特定类型化合物推广应用于其他需要大量信息表示和自动推导的化学物质虚拟筛选。

**1.1.3 基于靶体(药物-靶标相互作用)** 体外实验预测药物-靶标相互作用(DTI)费用高、耗时长,而机器学习算法将化学结构和靶点蛋白特征及其序列相结合进行预测,通过化合物与靶标之间结合亲和力的计算进行预测,能够大量减少实验室实验的数量,大大提高发现先导化合物的概率<sup>[20, 30-33]</sup>。

Ye 等<sup>[34]</sup>提出“AdvB-DTI”模型,模型中药物和靶标表达谱的特征通过矩阵分解与对抗性贝叶

斯个性化排序相关联, 根据已知的药物 - 靶标关系, 生成一组三元偏序关系, 使用对抗性贝叶斯个性化排序, 利用这些偏序关系训练药物和靶点的潜在因子矩阵, 并根据得分排名对 DTI 进行预测。由于此模型将流程与扰动因子和双重相抵正则相结合, 使模型更稳定, 训练结构也更加准确。Wang 等<sup>[35]</sup>提出“DLGRMC”模型, 是一种有效的对偶拉普拉斯图正则化矩阵补全计算模型。此模型将药物与靶标相互作用的预测转化为矩阵补全问题, 在矩阵补全过程中进行药物与靶标相互作用得分预测, 并且模型中充分利用率药物化学结构相似性和靶标基因组序列相似性, 即此模型对化学结构相似的药物和基因组序列相似的靶标相互作用具有预测作用。Yuan 等<sup>[36]</sup>提出“DrugE-Rank”模型, 将机器学习基于特征的方法和基于相似性的方法优势结合, 提高了预测性能。此模型将药物靶标相互作用的预测建模为多标签分类任务, 然后基于特征的机器学习方法将多标签分类, 转化成标签排名, 通过基于相似性的机器学习方法输出药物和靶标的特征作为学习排名的输入, 得到药物靶标相互作用结果预测。

## 1.2 机器学习应用于先导化合物的发现与优化

先导化合物是通过各种途径和手段得到的具有某种生物活性和化学结构的化合物, 用于进一步结构改造和修饰, 是现代新药研究的出发点。在新药研究过程中, 通过化合物活性筛选获得具有生物活性的先导化合物是创新药物研究的基础。先导化合物的生成和优化可以通过联合计算和实验研究实现, 使用从头设计和虚拟筛选两种方法<sup>[37]</sup>。目前已有计算机技术能够根据靶标蛋白结构直接构建与目标口袋结合的化合物, 并且已有大量化合物数据库可供使用, 通过虚拟筛选确定有价值的化合物, 对其进行修饰, 以获得具有生物活性的化合物。

**1.2.1 药物从头设计** 药物从头设计是一种基于结构的药物设计方法, 目标是提出具有预期药理活性和性质的新分子结构。常规药物分子从头设计包括基于结构和基于配体的设计方法, 而随着计算机辅助药物设计的发展, 人工智能算法、药物从头设计逐渐受到广泛关注<sup>[38-39]</sup>。

Wei 等<sup>[40]</sup>提出 CoV\_FB3D 方法, 为识别与靶标蛋白最匹配的共价化合物提供了一种机器学习方法。该方法基于靶标蛋白结合口袋中具有最大药效团特征的药物片段位置, 对有效的共价命中进行基于结构的计算机组装, 可采用综合评分策略评估每

种化合物的可合成性。Friedrich 等<sup>[41]</sup>提出一种能够合成天然产物模拟物的靶标识别和从头设计策略, 在快速的设计 - 制造 - 测试 - 分析周期中将自动化的、基于规则的分子构建与机器学习和实验验证相结合, 并用于目标分子预测。Morris 等<sup>[42]</sup>提出一种新的机器学习算法模型, 以自动生成具有化学多样性、合成可及性和生物活性的化合物为目标, 主要包括化合物优先排序和化学空间探索两部分, 且通过实验证明该模型在化合物扩展中具有实用性。

**1.2.2 药物重定位** 传统药物发现中通常使用一种“单靶点 - 单药物 - 特定疾病”的观念, 但这种观念存在一定局限性, 只能开发一种药物对某种疾病的单一靶点进行调节, 而具有次级靶点的药物可能会导致不良反应的发生, 也可能是新的疾病治疗的机会<sup>[43]</sup>。因此, 药物重定位对于药物研发是十分必要的。机器学习算法可将不同的靶点信息、小分子化合物特征以及疾病相关基因序列等信息相结合, 通过模拟计算打分进行药物重定位。

Emon 等<sup>[44]</sup>基于真实基因表达特征和 GWAS 数据提出“PS4DR”模型, 能够将多模式数据与不同数据库的路径信息集成, 预测不同疾病中药物重新定位的自动化工作流程。此模型通过 GWAS 数据筛选疾病和药物基因表达特征, 随后通过筛选的数据集进行路径分析, 并对疾病和药物特征进行计算, 最后通过计算疾病与药物路径特征的关联分数来执行反相关分析, 以确定不同疾病的药物优先顺序。Rodriguez 等<sup>[45]</sup>提出一种机器学习框架“DRIAD”, 用于评估阿尔兹海默病和能够通过基因列表描述的任何生物过程之间的潜在关系。与传统方法相比, 此框架不需要在整个基因空间的基础上构建模型, 在模型训练和预测期评估之前通过过滤转录组空间中与药物相关的基因获得有限的特征集, 直接、无差别地量化药物作用和阿尔兹海默病进展之间的关系。Zhao 等<sup>[46]</sup>提出一种机器学习工作流程, 通过基于药物表达谱预测特定疾病的适应症, 进行药物发现和重定位。此模型只考虑药物适应症和药物诱导的转录变化, 因此适用于任何可获得表达谱的化合物或药物。此外, 模型中对药物转录组的使用避免了明确靶点和了解作用机制, 甚至可将模型扩展到化学成分混合的药物中, 如中药。但此模型不包括的已知药物靶点、药物化学性质等有待改进。

## 1.3 机器学习应用于候选药物的确定

候选药物是指先导化合物经过结构修饰后得

到的化合物, 此类化合物的活性、安全性、药动学性质、选择性等并不确定, 是需要临床研究以确定其性质和修饰方案的化合物。在药物再利用过程中, 药物和候选药物都可与药物开发平台合作, 生成药物或候选药物的可行靶标谱以便对化合物进行筛选, 将药物和候选药物应用于新的靶点<sup>[47-49]</sup>。

大数据时代, 候选药物的数量集越来越大, 将机器学习算法应用于候选药物的确定, 能够有效提高药物研发效率, 且部分已报道算法模型还可应用于不同类别疾病模型的候选药物确定, 甚至可用于活性物质筛选相关的其他领域。

Theodoris 等<sup>[50]</sup>开发了识别广泛纠正人类多能干细胞疾病候选药物的一种机器学习方法。该方法进行了人类疾病相关的诱导多能干细胞的药物筛选, 且在人原代细胞和小鼠模型中进行了验证, 最终确定了靶向人类心脏病的潜在候选治疗分子。Oliveira 等<sup>[51]</sup>使用分子建模技术从药物数据库中获得对抗 SARS-CoV-2 的候选药物。在此研究中, 通过将分离的 S-蛋白溶解在水中进行分子动力学模拟, 观察到蛋白受体结合域构象转变, 且转变后的受体结合域更容易接触到溶剂和可能的药物。研究从平和的分子动力学结构出发, 通过对接计算对美国食品药品监督管理局 (FDA) 批准的药物库进行虚拟筛选。Margulis 等<sup>[52]</sup>提出机器学习方法模型“BitterIntense”, 此模型依据化合物的化学结构计算出描述符, 将化合物分为“非常苦”和“不很苦”两类, 对药物适口性进行预测, 且在测试集中具有 80% 以上的准确率。随后对“非常苦”化合物在数据集中进行评估, 以说明其性质与可能的毒性和疗效的关系, 以便为候选药物的确定提供参考。

#### 1.4 机器学习应用于成药性优化

成药性是先导化合物的优化和候选药物的目标, 指具有药理作用、能够进入 I 期临床并且具有

适宜药动学性质和安全性的药物性质, 成药性评估和优化是药物开发的重要过程<sup>[53-55]</sup>。目前, 成药性主要是对药物有效性、药动学特性和安全性进行评价。大数据和人工智能技术对药物安全性评价和改进提供了机会, 已有研究表明, 人工智能技术在药物安全评价中发挥重要作用, 能够通过评估药物各种特征来预测药物毒性<sup>[56]</sup>。机器学习算法用于药物成药性优化可从药物有效性、药动学特性和药物安全性评价等方面进行预测, 减少了药物设计和优化过程中常见的人为错误和偏差, 减少了测试所需的候选化合物的数量, 将候选化合物的测试时间大大缩短, 且使疾病生物学的重现比体外分析更有效<sup>[57]</sup>。

Chen 等<sup>[58]</sup>提出使用计算机辅助技术结合药物在大鼠体内的药动学、绝对生物利用度和组织分布来预测药物的成药性, 是一种活性化合物的快速筛选和初步评价的方法。首先利用分子对接技术检测药物与靶点的结合部位, 然后利用 ACD/Percepta 软件根据定量构效关系预测药效性, 最后通过高效液相色谱法测定药物在大鼠体内的药动学和组织分布。Zhang 等<sup>[59]</sup>利用朴素贝叶斯、支持向量机、递归划分、k 邻近、C4.5 决策树、随机森林和 Adboost 7 种机器学习方法建立了用于发育毒性预测的二进制分类模型。通过实验验证, 显示朴素贝叶斯分类器的预测性能和稳定性最好, 可用于药物开发中对药物化学发育毒性的预测。Zhou 等<sup>[60]</sup>提出一种基于多标签增强型随机森林的“MEDICASCY”机器学习模型。此模型在给定小分子结构的情况下, 即可对药物分子的不良反应、适应症、疗效和作用方式进行预测, 与许多已有方法模型的预测效果相当, 甚至更好, 且通过实验验证, 此模型可用于筛选小分子库中不良反应较小, 且适应症概率更高的药物, 为化合物成药性优化提供参考。

可使用机器学习模型信息见表 1。

表 1 可使用机器学习模型

Table 1 Available machine learning models

名称	网址	功能
eFindSite <sup>[23]</sup>	<a href="https://github.com/michal-brylinski/efindsite">https://github.com/michal-brylinski/efindsite</a>	蛋白质可药性预测
PharmMaker <sup>[25]</sup>	<a href="http://prody.csb.pitt.edu/pharmmaker">http://prody.csb.pitt.edu/pharmmaker</a>	构建药效团模型
DrugE-Rank <sup>[36]</sup>	<a href="http://datamining-iip.fudan.edu.cn/service/DrugE-Rank">http://datamining-iip.fudan.edu.cn/service/DrugE-Rank</a>	药物靶标相互作用预测
Cov_FB3D <sup>[40]</sup>	<a href="http://202.114.32.71:10280/wandox/Cov_FB3D/home.html">http://202.114.32.71:10280/wandox/Cov_FB3D/home.html</a>	化合物可合成性评估
PS4DR <sup>[44]</sup>	<a href="https://github.com/ps4dr/ps4dr">https://github.com/ps4dr/ps4dr</a>	不同疾病药物优先顺序预测
DRIAD <sup>[45]</sup>	<a href="https://labsyspharm.shinyapps.io/DRIAD/">https://labsyspharm.shinyapps.io/DRIAD/</a>	药物作用与阿尔兹海默病进展关系预测
MEDICASCY <sup>[60]</sup>	<a href="http://pwp.gatech.edu/cssb/MEDICASCY">http://pwp.gatech.edu/cssb/MEDICASCY</a>	可获得表达谱的化合物和药物重定位预测

## 2 深度学习在药物研发中的应用

深度学习是传统机器学习的延伸发展,也称为人工神经网络,是由连接的人工神经元组成的网络系统,从而模仿人类的中枢神经系统,其算法包括卷积神经网络、递归神经网络、深度神经网络、循环神经网络和自动编码器网络等<sup>[61-62]</sup>。

### 2.1 深度学习应用于药物靶标发现

**2.1.1 基于靶标(蛋白质可药性)** 靶标蛋白质具有相对分子质量大、分子结构复杂、种类极其多样和功能极为重要等特点,机器学习能够解释蛋白质序列及其同源性如何控制残基间的接触和结构组织<sup>[63]</sup>。而深度学习在生物信息学基础上对大量生物数据分类、分析,从而完成蛋白质靶标预测<sup>[64]</sup>。将深度学习与药物靶标发现相结合,可有效提高靶标发现效率、准确性和有效性。

Littmann 等<sup>[65]</sup>提出一种用于预测蛋白质残基是否与金属离子、核酸或小分子结合的深度学习模型“bindEmbed21”,由 bindEmbed21DL 和 bindEmbed21HBI 两个组件组成。该模型从预训练的蛋白质语言模型输入蛋白质表示,进而进行蛋白质残基与小分子结合预测,从而对所有可用的蛋白质序列进行快速预测。Kandel 等<sup>[66]</sup>提出了一个深度学习模型“PUResNet”和一种基于结构相似性的新型数据清理过程,用于预测蛋白质-配体结合位点。该模型主干为 ResNet 架构,包含编码器和解码器两个模块,两个模块间存在用于解决梯度消失问题的跳跃连接。经实验验证,在使用距离、体积和比例指标评估两个独立测试集 Coach420 和 BU48 时,取得了比现有方法 kalasanty 更好更合理的性能。Kozlovskii 等<sup>[67]</sup>提出了一种深度学习模型“BiteNet(结合位点神经网络)”,适用于蛋白质结合位点的大规模时空识别。该模型通过计算机视觉方法进行物体检测,将蛋白质的三维结构表示为具有对应于原子密度通道的 3D 图像,通过对构象集合的大规模分析来探索蛋白质特性,从而检测到相关结合位点,并用于基于结构的药物设计。

**2.1.2 基于配体** 深度学习算法能够从现有化合物中提取更完善的特征,并根据提取的特征设计新的化合物<sup>[68]</sup>,也对小分子配体的特征进行学习,建立数据集并开发数据预测模型,从而以配体为基础对药物-靶点相互作用进行智能预测。

Qiang 等<sup>[69]</sup>提出了应用迁移学习方法的深度学习算法“多层感知器(MLP)”,用于天然产物靶

标预测。该模型在去除天然产物信息的 ChEMBL 数据集上训练深度学习模型,借助数据集的知识学习结构和靶标之间的关系,随后使用具有更高效率的天然产物数据集对其进行微调,从而预测不同化合物的靶标,并调整其参数,以辅助基于天然产物结构的先导化合物发现。Gonczarek 等<sup>[70]</sup>提出 1 种预测蛋白质-分子对的结合能力深度学习框架,可用于基于结构的虚拟筛选,通过对分子应用可学习的原子卷积和 softmax 操作生成蛋白质和小分子指纹。将指纹进行进一步非线性变换,计算其内积,并用于预测蛋白质和小分子的结合潜力。同时,研究者提出建立在 PDBBind、DUD-E 和 MUV 数据上的新的基准数据集,为该深度学习模型识别目标蛋白配体提供参考。Khurana 等<sup>[71]</sup>提出一种用于蛋白质溶解度预测的深度学习模型“DeepSol”。该模型通过卷积神经网络识别蛋白质序列中的 k-mer 结构和蛋白质溶解度之间的关系,通过从原始输入序列中提取判别特征优化模型,提高溶解度预测,从而对药物研究和生产提供一定的参考。

**2.1.3 基于靶体(药物-靶点相互作用)** 传统机器学习方法往往将药物和靶标编码器表述为两个独立的模块,而不考虑它们之间的关系<sup>[72]</sup>,而深度学习算法模型可通过使用已知的药物-靶标相互作用信息对潜在的药物-靶标相互作用进行预测,无需额外的化学机构、蛋白质结构或序列等即可有效提高预测效率和精确度,大大加快了识别药物与靶蛋白相互作用的过程<sup>[73-74]</sup>。

Masoudi-Sobhanzadeh 等<sup>[75]</sup>提出“Trader”模型,使用此模型设计并训练了一个多层人工神经网络来对药物-靶标相互作用进行预测。经过准确度、敏感度、特异度和精密度等检测,其结果比以往算法更优。Liu 等<sup>[76]</sup>提出一种基于多信息加权融合的药物靶点相互作用预测方法,此方法将交互关系未知的样本视为未标记样本,筛选出可能具有交互作用但未经实验验证的样本,并根据筛选结果修改原始数据集。经过试验验证,发现该加权融合方法更加合理,提高了筛选结果的有效性和可靠性。但该方法存在一定局限性,在样本数较多的数据集中表现良好,但样本数较少的数据集中泛化能力不足。Chen 等<sup>[77]</sup>开发了一种药物-靶标相互作用预测系统,该系统使用了具有注意力机制和注意力双向长短期记忆(BiLSTM)图神经网络的端到端表示学习进行预测,且通过一种来自转换器预训练方法的

双向编码器从蛋白质序列中提取子结构特征，病因注入一种局部广度优先搜索模型，从分子图中学习子图信息。此方法在 Human Dataset、DUD-E Dataset 和 MUV target proteins 3 个数据集上均具有较好的预测性能，可用于筛选特定蛋白质的潜在药物。

## 2.2 深度学习应用于先导化合物的发现与优化

**2.2.1 药物从头设计** 随着高通量筛选和虚拟筛选技术的使用，基于配体的深度学习药物设计模型已经取得了一定的成功，但存在数据不足的问题，利用深度学习算法和分子建模方法对药物进行从头设计提供了一个有效的解决途径<sup>[78-79]</sup>。在目标靶点已知的情况下，深度学习模型可以根据不同的使用需求建模，此类模型大多通过长短期记忆网络进行循环训练和微调，以完全数据驱动的方式进行训练，并逐渐获得与靶点结合效率更高的药物分子，同时不需要大量的靶点配体数据支撑<sup>[80-81]</sup>。

Grisoni 等<sup>[82]</sup>提出深度学习自动化分子设计模型“设计-制造-测试-分析框架”，该模型基于长短期记忆细胞的循环神经网络用于生成对选定大分子靶标具有生物活性和可在微流体合成平台上合成的化合物，实验表明该模型捕获所需分子特性的能力以及支持自动化的潜力较好。Arshia 等<sup>[83]</sup>提出一种药物从头设计的深度学习方法，通过对长短期记忆网络进行训练和微调，生成随着训练逐渐增加结合能的配体。此外，该模型能够通过对接模拟的聚类分析确定排名靠前的代表性结构，并进行分子动力学模拟，然后进行分子力学泊松-玻尔兹曼表面积分析，验证得到的化合物分子的相互作用和结合亲和力。Popova 等<sup>[84]</sup>提出用于药物分子的从头设计模型“ReLeaSE (Reinforcement Learning for Structural Evolution)”。该模型在深度学习和强化学习方法的基础上集成了生成式和预测式两个深度神经网络，二者分别进行监督学习算法训练，随后都与强化学习方法联合训练，以便产生的化学分子结构具有所需的物理、化学或生物学特性。此模型具有不需要手动输入分子特征集的优点。除此之外，深度学习还能够应用于双靶点配体的从头生成。Lu 等<sup>[85]</sup>建立了深度学习计算框架“双靶点配体生成网络”，该模型通过对抗训练和强化学习，将基于序列的简化分子输入线输入系统生成器作为探索化学空间的随机策略，使用两个判别器鼓励生成属于两个生物活性化合物分别交叉点的分子。实验表明，该模型可以在多个基于结构的指标中生成与两

个生物活性数据集高度相似的新化合物，且与各种最先进的多目标分子生成模型具有相当的性能。

**2.2.2 药物重定位** 采用现有的药物重新定位方法进行捕捉高度非线性的、不同种类的网络结构具有一定的难度，而基于深度学习的大规模、异质生物网络算法为药物重定位提供了新的机遇<sup>[86]</sup>。

已报道的深度学习模型具有将药物与疾病相关联并对小分子化合物进行虚拟筛选的功能，从而将已有药物应用于新的靶点，适用于化学合成药物的同时也可应用于天然产物再利用。Liu 等<sup>[87]</sup>提出一个高通量的药物再利用计算模型，遵循了随机临床试验设计方案，且通过计算筛选重定位药物，由于实际数据具有时间序列和混杂变量的特征，应用深度学习和因果推理方法即长短期记忆和治疗加权逆概率来控制实际数据中的混杂因素，并系统估计药物对不同疾病的影响。与临床前方法相比，此模型具有转化问题更少、更稳定的优势。Yi 等<sup>[88]</sup>开发了一种深度门控循环单元模型“DDIPreD”，使用综合相似性度量和高斯交互轮廓核和门控循环神经网络预测潜在的药物-疾病关联关系，发现现有的药物新适应症，从而促进药物研发进程。且实验结果证明，该模型在两个基准数据集数据集上取得显著性能，能够有效预测药物新适应症或疾病新疗法，加速药物重新定位和相关药物研究发现。Xu 等<sup>[89]</sup>提出一种基于深度学习的虚拟筛选系统，该模型在相关数据集中获取一个天然产物分子，使用神经网络模型预测该分子的溶剂分解和氧化产物，随后将预测产物与同一生物来源的其他天然产物分子进行匹配，匹配成功后即可将天然产物和预测产物标记为潜在药物分子，从而对天然产物、人工合成分子的药物进行预测发现。

## 2.3 深度学习应用于候选药物的确定

深度学习模型能够独立学习复杂的特征信息，减少甚至避免人工输入过程，标准化、可靠的大型数据集能够进行较准确地预测。利用神经网络对数据库中的分子进行学习，将原始分子转换为分子指纹图谱等信息，可大大提高人工智能算法的分子预测性能。因此，将有效的深度学习模型用于候选药物确定，可以在很大程度上提高药物发现的效率。

Puentes 等<sup>[90]</sup>提出一种用于预测分子与可能蛋白质靶点受体结合的深度学习框架“PharmaNet”，利用循环神经网络在大型数据库中进行主动分子预测，进行候选药物筛选。该算法

主要包括 SMILES 将分子转换为原始分子图像、卷积编码器处理数据和循环神经网络分析的指纹分子图像 3 个阶段。该模型的预测性能提升到 65.5%，超过目前已有大部分模型。Woo 等<sup>[91]</sup>提出“DeepCOP”深度学习模型，该模型结合分子指纹描述符和基因描述符对预测 LINCS 数据库中收集的差异基因调控端点的深度神经网络进行训练。此模型应用于内部训练集时预测效果较好，但应用于外部数据集时效果不理想，可依据更丰富的外部数据和更标准化和可控的实验条件改进，从而用于筛选不断增长的大量可用化学物质库，寻找具有所需基因调节特性并可以操纵细胞通路的候选药物。Arshadi 等<sup>[92]</sup>提出深度学习模型“DeepMalaria”，能够使用 SMILES 预测化合物抗恶性疟原虫抑制特性。该模型在公开数据集上进行训练，使用大量不相关数据的迁移学习对进行补充训练，使用自动特征提取学习分子内模式，并推广到新的、未知的数据集中，在独立的大环测试数据集上进行测试，从而确定新的候选药物。

#### 2.4 深度学习应用于成药性优化

深度学习算法可以将疾病因子、患者情况、药理学参数和候选药物特征等数据转化成图谱，并对候选药物进行识别，将疾病和药物分子匹配、优化，

进行疗效和不良反应预测，从而使化合物的成药性提高<sup>[16, 93-95]</sup>。

Zhu 等<sup>[96]</sup>提出基于深度学习的疗效预测系统“DLEPS”，可以通过输入患病状态下基因表达谱对候选药物进行识别，只需要基因特征即可预测高度多样化和复杂疾病的候选分子，适用于数以万计的功能基因组学研究，可对所有候选疾病治疗化合物进行疗效预测，还可以深入了解致病机制和易受攻击的靶蛋白和途径。Anastopoulos 等<sup>[97]</sup>提出一种将患者特定的人口统计学、临床和遗传特征与药物结构相结合的模型，为可扩展的图卷积方法，能够整合典型患者正在服用药物的分子效应，从而预测药物的不良反应。且该模型在预测 UK Biobank 数据集任务中明显优于标准机器学习方法，有很大可能用于现实中不同人群中药物的个体化毒性，对其成药性提供一定的参考。Ye 等<sup>[98]</sup>建立了一个可用于药理学参数预测的算法，该算法包含 4 个关键人体药理学参数数据，并引入大型生物活性数据集，使用集成迁移学习和多任务学习方法训练深度神经网络。与其他机器学习方法相比，该算法可以自动从原始数据中提取关键特征或分子描述符，将原始数据转换为更高级别的特征，且具有更高的准确性和泛化能力。可使用深度学习模型信息见表 2。

表 2 可使用深度学习模型  
Table 2 Available deep learning models

名称	网址	功能
bindEmbed21 <sup>[65]</sup>	<a href="https://github.com/Rostlab/bindPredict">https://github.com/Rostlab/bindPredict</a>	对可用蛋白序列进行快速预测
PUResNet <sup>[66]</sup>	<a href="https://github.com/jivankandel/PUResNet">https://github.com/jivankandel/PUResNet</a>	蛋白质 - 配体结合位点预测
BiteNet <sup>[67]</sup>	<a href="https://sites.skoltech.ru/imolecule/tools/bitenet">https://sites.skoltech.ru/imolecule/tools/bitenet</a>	蛋白质结合位点识别
DeepSol <sup>[71]</sup>	<a href="https://zenodo.org/record/1162886">https://zenodo.org/record/1162886</a>	蛋白质溶解度预测
Trader <sup>[75]</sup>	<a href="https://github.com/LBBSof/Trader">https://github.com/LBBSof/Trader</a>	药物 - 靶标相互作用预测
— <sup>[77]</sup>	<a href="https://github.com/Fitnessnp/DeepEmbedding-DTI">https://github.com/Fitnessnp/DeepEmbedding-DTI</a>	药物 - 靶标相互作用预测
— <sup>[82]</sup>	<a href="https://github.com/ETHmodlab/virtual_libraries">https://github.com/ETHmodlab/virtual_libraries</a>	有活性并可合成的化合物预测
ReLeaSE <sup>[84]</sup>	<a href="https://github.com/isayev/ReLeaSE">https://github.com/isayev/ReLeaSE</a>	药物分子从头设计
— <sup>[87]</sup>	<a href="https://github.com/ruoqi-liu/DeepIPWor">https://github.com/ruoqi-liu/DeepIPWor</a>	筛选重定位药物
DDIPred <sup>[88]</sup>	<a href="https://github.com/haichengyi/DDIPred">https://github.com/haichengyi/DDIPred</a>	潜在的药物 - 疾病关联预测
— <sup>[89]</sup>	<a href="http://www.mdpi.com/2218-273X/10/11/1486/s1">http://www.mdpi.com/2218-273X/10/11/1486/s1</a>	天然产物与人工合成分子药物预测
PharmaNet <sup>[90]</sup>	<a href="https://github.com/BCV-Uniandes/PharmaNet">https://github.com/BCV-Uniandes/PharmaNet</a>	分子与可能蛋白质靶点受体集合预测
DeepCOP <sup>[91]</sup>	<a href="https://github.com/godwinwoo/DeepCOP">https://github.com/godwinwoo/DeepCOP</a>	候选药物预测
DeepMalaria <sup>[92]</sup>	<a href="http://www.transilico.com">www.transilico.com</a>	候选药物预测
DLEPS <sup>[96]</sup>	<a href="https://github.com/kekegg/DLEPS">https://github.com/kekegg/DLEPS</a>	候选化合物疗效、致病机制、靶蛋白、途径预测

### 3 结语与展望

随着大数据的广泛应用和计算机技术不断发展, 药物研发领域也不断走向人工智能。根据已有文献报道, 人工智能技术在药物研发的各个环节均有应用, 虽然大多数机器学习和深度学习算法的应用并不尽善尽美, 但也都在某方面甚至几个方面具有突出优势, 且随着算法的不断优化, 人工智能模型预测结果也更加准确, 甚至能够预测出训练集之外的衍生数据。将人工智能技术与药物研发相结合是一个不断发展的过程, 从药物发现到药效评价的每一个环节都可以开发相应的技术模型, 人工智能技术将可能成为未来药物研发的一个重要途径, 能够广泛应用于化学药物、天然产物以及中药的开发, 同时药物研发方法也迈上新的台阶。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] Hermann E, Hermann G, Tremblay J C. Ethical artificial intelligence in chemical research and development: A dual advantage for sustainability [J]. *Sci Eng Ethics*, 2021, 27(4): 45.
- [2] Zhong F, Xing J, Li X, et al. Artificial intelligence in drug design [J]. *Sci China Life Sci*, 2018, 61(10): 1191-1204.
- [3] Jimenez-Luna J, Grisoni F, Weskamp N, et al. Artificial intelligence in drug discovery: Recent advances and future perspectives [J]. *Expert Opin Drug Discov*, 2021, 16(9): 949-959.
- [4] Cova T, Pais A. Deep learning for deep chemistry: Optimizing the prediction of chemical patterns [J]. *Front Chem*, 2019, 7: 809.
- [5] Hessler G, Baringhaus K H. Artificial intelligence in drug design [J]. *Molecules*, 2018, 23(10): 2520.
- [6] Zhu H. Big data and artificial intelligence modeling for drug discovery [J]. *Ann Rev Pharmacol Toxicol*, 2020, 60: 573-589.
- [7] Gupta R, Srivastava D, Sahu M, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery [J]. *Mol Divers*, 2021, 25(3): 1315-1360.
- [8] Yang X, Wang Y, Byrne R, et al. Concepts of artificial intelligence for computer-assisted drug discovery [J]. *Chem Rev*, 2019, 119(18): 10520-10594.
- [9] Malik A, Patel P, Ehsan L, et al. Ten simple rules for engaging with artificial intelligence in biomedicine [J]. *PLoS Comput Biol*, 2021, 17(2): e1008531.
- [10] Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development [J]. *Nat Rev Drug Discov*, 2019, 18(6): 463-477.
- [11] Lv H, Shi L, Berkenpas J W, et al. Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design [J]. *Brief Bioinform*, 2021, 22(6): bbab320.
- [12] David L, Arus-Pous J, Karlsson J, et al. Applications of deep-learning in exploiting large-scale and heterogeneous compound data in industrial pharmaceutical research [J]. *Front Pharmacol*, 2019, 10: 1303.
- [13] Patel L, Shukla T, Huang X, et al. Machine learning methods in drug discovery [J]. *Molecules*, 2020, 25(22): 5277.
- [14] Nag S, Baidya A T K, Mandal A, et al. Deep learning tools for advancing drug discovery and development [J]. *3 Biotech*, 2022, 12(5): 110.
- [15] Han K, Zhang L, Wang M, et al. Prediction methods of herbal compounds in Chinese medicinal herbs [J]. *Molecules*, 2018, 23(9): 2303.
- [16] Kimber T B, Chen Y, Volkamer A. Deep learning in virtual screening: Recent applications and developments [J]. *Int J Mol Sci*, 2021, 22(9): 4435.
- [17] Hu Y, Lu Y, Wang S, et al. Application of machine learning approaches for the design and study of anticancer drugs [J]. *Curr Drug Targets*, 2019, 20(5): 488-500.
- [18] 刘润哲, 宋俊科, 刘艾林, 等. 人工智能在基于配体和受体结构的药物筛选中的应用进展 [J]. *药学报*, 2021, 56(8): 2136-2145.
- [19] Vatansever S, Schlessinger A, Wacker D, et al. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions [J]. *Med Res Rev*, 2021, 41(3): 1427-1473.
- [20] Chen R, Liu X, Jin S, et al. Machine learning for drug-target interaction prediction [J]. *Molecules*, 2018, 23(9): 2208.
- [21] Eder J, Herrling P L. Trends in modern drug discovery [J]. *Handb Exp Pharmacol*, 2016, 232: 3-22.
- [22] Paananen J, Fortino V. An omics perspective on drug target discovery platforms [J]. *Brief Bioinform*, 2020, 21(6): 1937-1953.
- [23] Kana O, Brylinski M. Elucidating the druggability of the human proteome with eFindSite [J]. *J Comput Aided Mol Des*, 2019, 33(5): 509-519.
- [24] Cavasotto C N, Lamas M S, Maggini J. Functional and druggability analysis of the SARS-CoV-2 proteome [J]. *Eur J Pharmacol*, 2021, 890: 173705.
- [25] Lee J Y, Krieger J M, Li H, et al. Pharmmaker: Pharmacophore modeling and hit identification based on druggability simulations [J]. *Protein Sci*, 2020, 29(1): 76-86.



- [26] Raschka S, Kaufman B. Machine learning and AI-based approaches for bioactive ligand discovery and GPCR-ligand recognition [J]. *Methods*, 2020, 180: 89-110.
- [27] Chaube S, Srinivasan S G, Rai B. Applied machine learning for predicting the lanthanide-ligand binding affinities [J]. *Sci Rep*, 2020, 10(1): 14322.
- [28] Da'adoosh B, Marcus D, Rayan A, et al. Discovering highly selective and diverse PPAR-delta agonists by ligand based machine learning and structural modeling [J]. *Sci Rep*, 2019, 9(1): 1106.
- [29] Sun Y, Zhou H, Zhu H, et al. Ligand-based virtual screening and inductive learning for identification of SIRT1 inhibitors in natural products [J]. *Sci Rep*, 2016, 6: 19312.
- [30] Zhao T, Hu Y, Valsdottir L R, et al. Identifying drug-target interactions based on graph convolutional network and deep neural network [J]. *Brief Bioinform*, 2021, 22(2): 2141-2150.
- [31] Zhou L, Li Z, Yang J, et al. Revealing drug-target interactions with computational models and algorithms [J]. *Molecules*, 2019, 24(9): 1714.
- [32] Zhao Q, Yang M, Cheng Z, et al. Biomedical data and deep learning computational models for predicting compound-protein relations [J]. *IEEE/ACM Trans Comput Biol Bioinform*, 2022, 19(4): 2092-2110.
- [33] D'Souza S, Prema K V, Balaji S. Machine learning models for drug-target interactions: current knowledge and future directions [J]. *Drug Discov Today*, 2020, 25(4): 748-756.
- [34] Ye Y, Wen Y, Zhang Z, et al. Drug-target interaction prediction based on adversarial Bayesian personalized ranking [J]. *Biomed Res Int*, 2021, 2021: 6690154.
- [35] Wang M, Tang C, Chen J. Drug-target interaction prediction via dual Laplacian graph regularized matrix completion [J]. *Biomed Res Int*, 2018, 2018: 1425608.
- [36] Yuan Q, Gao J, Wu D, et al. DrugE-rank: Improving drug-target interaction prediction of new candidate drugs or targets by ensemble learning to rank [J]. *Bioinformatics*, 2016, 32(12): i18-i27.
- [37] Jorgensen W L. Efficient drug lead discovery and optimization [J]. *Acc Chem Res*, 2009, 42(6): 724-733.
- [38] Chavez-Hernandez A L, Juarez-Mercado K E, Saldivar-Gonzalez F I, et al. Towards the De Novo design of HIV-1 protease inhibitors based on natural products [J]. *Biomolecules*, 2021, 11(12): 1805.
- [39] Mouchlis V D, Afantitis A, Serra A, et al. Advances in de Novo drug design: From conventional to machine learning methods [J]. *Int J Mol Sci*, 2021, 22(4): 1676.
- [40] Wei L, Wen W, Rao L, et al. Cov\_FB3D: A De Novo covalent drug design protocol integrating the BA-SAMP strategy and machine-learning-based synthetic tractability evaluation [J]. *J Chem Inf Model*, 2020, 60(9): 4388-4402.
- [41] Friedrich L, Cingolani G, Ko Y H, et al. Learning from nature: From a marine natural product to synthetic cyclooxygenase-1 inhibitors by automated De Novo design [J]. *Adv Sci (Weinh)*, 2021, 8(16): e2100832.
- [42] Morris A, McCorkindale W, Consortium T C M, et al. Discovery of SARS-CoV-2 main protease inhibitors using a synthesis-directed de novo design model [J]. *Chem Commun (Camb)*, 2021, 57(48): 5909-5912.
- [43] Galati S, Di Stefano M, Martinelli E, et al. Recent advances in in silico target fishing [J]. *Molecules*, 2021, 26(17): 5124.
- [44] Emon M A, Domingo-Fernandez D, Hoyt C T, et al. PS4DR: A multimodal workflow for identification and prioritization of drugs based on pathway signatures [J]. *BMC Bioinformatics*, 2020, 21(1): 231.
- [45] Rodriguez S, Hug C, Todorov P, et al. Machine learning identifies candidates for drug repurposing in Alzheimer's disease [J]. *Nat Commun*, 2021, 12(1): 1033.
- [46] Zhao K, So H C. Drug repositioning for schizophrenia and depression/anxiety disorders: A machine learning approach leveraging expression data [J]. *IEEE J Biomed Health Inform*, 2019, 23(3): 1304-1315.
- [47] Allarakhia M. Open-source approaches for the repurposing of existing or failed candidate drugs: learning from and applying the lessons across diseases [J]. *Drug Des Devel Ther*, 2013, 7: 753-766.
- [48] Parvathaneni V, Kulkarni N S, Muth A, et al. Drug repurposing: A promising tool to accelerate the drug discovery process [J]. *Drug Discov Today*, 2019, 24(10): 2076-2085.
- [49] 李敬, 向毅, 徐淑静, 等. 冠状病毒抑制剂研究的药物化学策略 [J]. *药学学报*, 2020, 55(4): 537-553.
- [50] Theodoris C V, Zhou P, Liu L, et al. Network-based screen in iPSC-derived cells reveals therapeutic candidate for heart valve disease [J]. *Science*, 2021, 371(6530): eabd0724.
- [51] de Oliveira O V, Rocha G B, Paluch A S, et al. Repurposing approved drugs as inhibitors of SARS-CoV-2 S-protein from molecular modeling and virtual screening [J]. *J Biomol Struct Dyn*, 2021, 39(11): 3924-3933.
- [52] Margulis E, Dagan-Wiener A, Ives R S, et al. Intense bitterness of molecules: Machine learning for expediting drug discovery [J]. *Comput Struct Biotechnol J*, 2021, 19: 568-576.
- [53] Vo A H, Van Vleet T R, Gupta R R, et al. An Overview of machine learning and big data for drug toxicity evaluation

- [J]. *Chem Res Toxicol*, 2020, 33(1): 20-37.
- [54] Romm E L, Tsigelny I F. Artificial intelligence in drug treatment [J]. *Ann Rev Pharmacol Toxicol*, 2020, 60: 353-369.
- [55] Muller C, Rabal O, Diaz Gonzalez C. Artificial intelligence, machine learning, and deep learning in real-life drug design cases [J]. *Methods Mol Biol*, 2022, 2390: 383-407.
- [56] Basile A O, Yahi A, Tatonetti N P. Artificial intelligence for drug toxicity and safety [J]. *Trends Pharmacol Sci*, 2019, 40(9): 624-635.
- [57] Thomford N E, Senthebane D A, Rowe A, *et al.* Natural products for drug discovery in the 21st century: Innovations for novel drug discovery [J]. *Int J Mol Sci*, 2018, 19(6): 1578.
- [58] Chen F, Yin X, Wang Y, *et al.* Pharmacokinetics, tissue distribution, and druggability prediction of the natural anticancer active compound cytosine *N*-isoflavones combined with computer simulation [J]. *Biol Pharm Bull*, 2020, 43(6): 976-984.
- [59] Zhang H, Mao J, Qi H Z, *et al.* In silico prediction of drug-induced developmental toxicity by using machine learning approaches [J]. *Mol Divers*, 2020, 24(4): 1281-1290.
- [60] Zhou H, Cao H, Matyunina L, *et al.* MEDICASCY: A machine learning approach for predicting small-molecule drug side effects, indications, efficacy, and modes of action [J]. *Mol Pharm*, 2020, 17(5): 1558-1574.
- [61] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [62] Jing Y, Bian Y, Hu Z, *et al.* Deep learning for drug design: An artificial intelligence paradigm for drug discovery in the big data era [J]. *AAPS J*, 2018, 20(3): 58.
- [63] Suh D, Lee J W, Choi S, *et al.* Recent applications of deep learning methods on evolution- and contact-based protein structure prediction [J]. *Int J Mol Sci*, 2021, 22(11): 6032.
- [64] Cui F, Zhang Z, Zou Q. Sequence representation approaches for sequence-based protein prediction tasks that use deep learning [J]. *Brief Funct Genomics*, 2021, 20(1): 61-73.
- [65] Littmann M, Heinzinger M, Dallago C, *et al.* Protein embeddings and deep learning predict binding residues for various ligand classes [J]. *Sci Rep*, 2021, 11(1): 23916.
- [66] Kandel J, Tayara H, Chong K T. PURESNet: Prediction of protein-ligand binding sites using deep residual neural network [J]. *J Cheminform*, 2021, 13(1): 65.
- [67] Kozlovskii I, Popov P. Spatiotemporal identification of druggable binding sites using deep learning [J]. *Commun Biol*, 2020, 3(1): 618.
- [68] Palazzesi F, Pozzan A. Deep learning applied to ligand-based De Novo drug design [J]. *Methods Mol Biol*, 2022, 2390: 273-299.
- [69] Qiang B, Lai J, Jin H, *et al.* Target prediction model for natural products using transfer learning [J]. *Int J Mol Sci*, 2021, 22(9): 4632.
- [70] Gonczarek A, Tomczak J M, Zareba S, *et al.* Interaction prediction in structure-based virtual screening using deep learning [J]. *Comput Biol Med*, 2018, 100: 253-258.
- [71] Khurana S, Rawi R, Kunji K, *et al.* DeepSol: A deep learning framework for sequence-based protein solubility prediction [J]. *Bioinformatics*, 2018, 34(15): 2605-2613.
- [72] Yang Z, Zhong W, Zhao L, *et al.* ML-DTI: Mutual learning mechanism for interpretable drug-target interaction prediction [J]. *J Phys Chem Lett*, 2021, 12(17): 4247-4261.
- [73] Wu Z, Li W, Liu G, *et al.* Network-based methods for prediction of drug-target interactions [J]. *Front Pharmacol*, 2018, 9: 1134.
- [74] Manoochehri H E, Nourani M. Drug-target interaction prediction using semi-bipartite graph model and deep learning [J]. *BMC Bioinformatics*, 2020, 21(Suppl 4): 248.
- [75] Masoudi-Sobhanzadeh Y, Omidi Y, Amanlou M, *et al.* Trader as a new optimization algorithm predicts drug-target interactions efficiently [J]. *Sci Rep*, 2019, 9(1): 9348.
- [76] Liu S, An J, Zhao J, *et al.* Drug-target interaction prediction based on multisource information weighted fusion [J]. *Contrast Media Mol Imaging*, 2021, 2021: 6044256.
- [77] Chen W, Chen G, Zhao L, *et al.* Predicting drug-target interactions with deep-embedding learning of graphs and sequences [J]. *J Phys Chem A*, 2021, 125(25): 5633-5642.
- [78] Krishnan S R, Bung N, Bulusu G, *et al.* Accelerating De Novo drug design against novel proteins using deep learning [J]. *J Chem Inf Model*, 2021, 61(2): 621-630.
- [79] Wang M, Wang Z, Sun H, *et al.* Deep learning approaches for de novo drug design: An overview [J]. *Curr Opin Struct Biol*, 2022, 72: 135-144.
- [80] Li Y, Pei J, Lai L. Structure-based de novo drug design using 3D deep generative models [J]. *Chem Sci*, 2021, 12(41): 13664-13675.
- [81] Carpenter K A, Cohen D S, Jarrell J T, *et al.* Deep learning and virtual drug screening [J]. *Future Med Chem*, 2018, 10(21): 2557-2567.
- [82] Grisoni F, Huisman B J H, Button A L, *et al.* Combining generative artificial intelligence and on-chip synthesis for de novo drug design [J]. *Sci Adv*, 2021, 7(24): eabg3338.
- [83] Arshia A H, Shadravan S, Solhjoo A, *et al.* De novo design of novel protease inhibitor candidates in the treatment of SARS-CoV-2 using deep learning, docking, and molecular

- dynamic simulations [J]. *Comput Biol Med*, 2021, 139: 104967.
- [84] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design [J]. *Sci Adv*, 2018, 4(7): eaap7885.
- [85] Lu F, Li M, Min X, *et al.* De novo generation of dual-target ligands using adversarial training and reinforcement learning [J]. *Brief Bioinform*, 2021, 22(6): bbab333.
- [86] Zeng X, Zhu S, Liu X, *et al.* deepDR: A network-based deep learning approach to in silico drug repositioning [J]. *Bioinformatics*, 2019, 35(24): 5191-5198.
- [87] Liu R, Wei L, Zhang P. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data [J]. *Nat Mach Intell*, 2021, 3(1): 68-75.
- [88] Yi H C, You Z H, Wang L, *et al.* In silico drug repositioning using deep learning and comprehensive similarity measures [J]. *BMC Bioinformatics*, 2021, 22(Suppl 3): 293.
- [89] Xu T, Chen W, Zhou J, *et al.* Virtual screening for reactive natural products and their probable artifacts of solvolysis and oxidation [J]. *Biomolecules*, 2020, 10(11): 1486.
- [90] Puentes P R, Valderrama N, Gonzalez C, *et al.* PharmaNet: Pharmaceutical discovery with deep recurrent neural networks [J]. *PLoS One*, 2021, 16(4): e0241728.
- [91] Woo G, Fernandez M, Hsing M, *et al.* DeepCOP: Deep learning-based approach to predict gene regulating effects of small molecules [J]. *Bioinformatics*, 2020, 36(3): 813-818.
- [92] Arshadi A K, Salem M, Collins J, *et al.* DeepMalaria: Artificial intelligence driven discovery of potent antiplasmodials [J]. *Front Pharmacol*, 2019, 10: 1526.
- [93] Caceres E L, Tudor M, Cheng A C. Deep learning approaches in predicting ADMET properties [J]. *Future Med Chem*, 2020, 12(22): 1995-1999.
- [94] Ekins S. The next era: Deep learning in pharmaceutical research [J]. *Pharm Res*, 2016, 33(11): 2594-2603.
- [95] Muzio G, O'Bray L, Borgwardt K. Biological network analysis with deep learning [J]. *Brief Bioinform*, 2021, 22(2): 1515-1530.
- [96] Zhu J, Wang J, Wang X, *et al.* Prediction of drug efficacy from transcriptional profiles with deep learning [J]. *Nat Biotechnol*, 2021, 39(11): 1444-1452.
- [97] Anastopoulos I N, Herczeg C K, Davis K N, *et al.* Multi-drug featurization and deep learning improve patient-specific predictions of adverse events [J]. *Int J Environ Res Public Health*, 2021, 18(5): 2600.
- [98] Ye Z, Yang Y, Li X, *et al.* An integrated transfer learning and multitask learning approach for pharmacokinetic parameter prediction [J]. *Mol Pharm*, 2019, 16(2): 533-541.

[责任编辑 解学星]