

## 药物发现过程中人工智能的应用研究进展

李双星<sup>1#</sup>, 李一昊<sup>2#</sup>, 林志<sup>1</sup>, 张 頔<sup>1</sup>, 杨艳伟<sup>1</sup>, 屈 哲<sup>1</sup>, 李言川<sup>2</sup>, 霍桂桃<sup>1\*</sup>, 吕建军<sup>2\*</sup>

1. 中国食品药品检定研究院, 国家药物安全评价监测中心, 北京市重点实验室, 北京 100176

2. 湖北天勤鑫圣生物科技有限公司, 湖北 武汉 430207

**摘要:** 人工智能(AI)和机器学习不仅使药物发现和开发实现了质的飞跃,而且帮助药物开发进程进入现代化。机器学习和深度学习算法已应用于药物发现各个阶段,如先导化合物的筛选、多肽合成及小分子药物的发现、最佳给药剂量的确定、类药化合物的设计和药物不良反应的预测、蛋白质间相互作用的预测、虚拟筛选效率的提高、定量构效关系(QSAR)建模和药物重新定位、理化性质和药物靶标亲和力的预测、化合物的结合预测和体内安全性分析、多靶点配体药物分子的设计以及临床试验的设计。简要综述了AI算法和传统化学相结合以提高药物发现的效率以及AI在药物发现过程中的应用研究进展,以期为AI应用于药物发现提供一定参考。

**关键词:** 人工智能; 药物发现; 药物开发; 机器学习; 定量构效关系; 药物靶标

中图分类号: R914.2 文献标志码: A 文章编号: 1674-6376(2023)09-2030-07

DOI: 10.7501/j.issn.1674-6376.2023.09.026

## Research progress in application of artificial intelligence during process of drug discovery

LI Shuangxing<sup>1</sup>, LI Yihao<sup>2</sup>, LIN Zhi<sup>1</sup>, ZHANG Di<sup>1</sup>, YANG Yanwei<sup>1</sup>, QU Zhe<sup>1</sup>, LI Yanchuan<sup>2</sup>, HUO Guitao<sup>1</sup>, LYU Jianjun<sup>2</sup>

1. Beijing Key Laboratory, National Center for safety Evaluation of Drugs, National Institutes for Food and Drug Control, Beijing 100176, China

2. Hubei Topgene Xinsheng Biotechnology Co., Ltd., Wuhan 430207, China

**Abstract:** Artificial intelligence (AI) and machine learning (ML) not only has made a qualitative leap in drug discovery and development, but also has helped the process of drug development enter modernization. Algorithms of ML and deep learning (DL) have been applied to various phases of drug discovery, such as screening of lead compounds, peptide synthesis and discovery of small molecular drugs, determination of the optimal dosage, design of drug-like compounds and prediction of adverse drug reaction (ADR), prediction of protein-protein interaction, improvement of the efficiency of virtual screening, quantitative structure-activity relationship (QSAR) modeling and drug repositioning, prediction of physicochemical properties and affinity of drug targets, prediction of drug binding and *in vivo* safety analysis, design of multiple target ligand drug molecules, as well as design of clinical trials. The research progress in the combination of AI algorithm and traditional chemistry to improve the efficiency of drug discovery and application of AI during the process of drug discovery was briefly reviewed, in order to provide some references for using AI for drug discovery in China.

**Key words:** artificial intelligence (AI); drug discovery; drug development; machine learning; quantitative structure-activity relationship (QSAR); drug targets

在过去的20年中,生物学家和化学家一直致力于开发一种高效和先进的药物发现和评价系统<sup>[1]</sup>,不仅可以高效率地发现靶向治疗药物,而且可以降低潜在药物不良反应的发生风险。为了减少新药

收稿日期: 2023-01-28

基金项目: 中国食品药品检定研究院学科带头人课题(2021X2)

#共同第一作者: 李双星(1995—),男,硕士,研究方向为药物非临床安全性评价。E-mail: lishuangxing@nifdc.org.cn

李一昊(1996—),男,硕士,研究方向为药物非临床安全性评价。E-mail: liyihao@topgenebio.com

\*共同通信作者: 霍桂桃(1978—),女,主任药师,研究方向为药物非临床安全性评价。E-mail: huoguitao@nifdc.org.cn

吕建军(1970—),男,主任药师,研究方向为药物非临床安全性评价。E-mail: lvjianjun@topgenebio.com

开发的时间并降低费用,科研人员将目光转向了基于计算机和大数据的方法,如虚拟筛选(VS)和分子对接,但这些方法具有准确性差和效率低等问题<sup>[2]</sup>。人工智能(AI),包括深度学习(DL)和机器学习(ML)算法,有望可以克服在药物设计和发现过程中遇到的诸多问题和困难。基于AI原理的计算机建模为化合物的识别和验证、靶标鉴定、多肽合成、药物毒性及理化性质的评价、药物监测、药物疗效以及药物重新定位提供了一个崭新和好的思路<sup>[3]</sup>。

2015年9月的谷歌搜索趋势显示,自ML引入后,AI是搜索最多的术语之一。部分学者将ML描述为主要的AI应用程序,另一些学者则认为ML是AI的1个子集<sup>[4-6]</sup>。Russell等<sup>[7]</sup>认为AI有7个亚类,分别是推理和问题解决、知识表示、规划和社会智能、感知、机器学习、机器人(运动和操纵)以及自然语言处理。简而言之,DL是ML的子集,ML是AI的子集,三者发展顺序依次是AI>ML>DL。

目前,有多种ML技术可以帮助科学家们在生物学大数据中寻找具有价值的特征、模式和结构。美国国立生物技术中心(NCBI)基因表达综合数据库(GEO)<sup>[8]</sup>、癌症基因组图谱(TCGA)和ArrayExpress<sup>[9]</sup>均是包含基因表达数据的大型数据库。通过分析基因表达特征,科学家们可以找出导致不同疾病的靶标基因。例如,利用ML方法和基因表达数据,van Jzendorn等<sup>[10]</sup>发现了罕见软组织肉瘤新的生物标志物和潜在的药物靶标。此外还有许多大型化学数据库可供科学家们使用,帮助其发现针对特定靶标的药物,例如PubChem就是一个可以免费访问的化学物质数据库,其中涵盖了各种化学结构的数据,包括生物、物理、化学和毒性特性。ChEMBL数据库包含许多生物活性化合物的吸收、分布、代谢和排泄(ADME)特征,以及靶标相互作用信息<sup>[11-12]</sup>。除此以外,还有诸如DrugBank<sup>[13]</sup>、LINCSL1000<sup>[14]</sup>和PDB等类似的数据库可供药物开发使用。

在制药行业,AI可以帮助解决经典化学中影响药物发现和开发的诸多问题。随着近10年信息技术进步和高性能计算机的快速发展,计算机辅助药物设计(CADD)中从ML到DL的一系列AI算法日益增加。目前,科学家们的目标是通过基于经典化学性质的ML算法改进药物的发现和开发过程,并使其具有较高的准确性和置信度。在过去的20年里建立了许多药物计算发现、定量构效关

系(QSAR)建模和自由能最小化方法等技术和工具<sup>[15]</sup>。Zhang等<sup>[16]</sup>对一种新型冠状病毒进行了计算机分析,筛选出对严重急性呼吸综合征(SARS)具有生物活性的化合物。随后,对这些化合物进行了吸收、分布、代谢、排泄及毒性(ADMET)和对接分析。结果表明有13种中药对新型冠状病毒有效。因此,传统化学导向的药物发现和开发概念与CADD相结合将为药物研发提供广阔的研究平台。但是,药物发现和开发过程中如何寻找合适且具有生物活性的药物分子是最难的一步。据不完全统计,约有90%的药物分子通常无法通过II期临床试验和监管机构的批准<sup>[17-20]</sup>。令人鼓舞的是,近10年来随着AI技术和多种算法的不断建立,以及生物学大数据的不断完善,药物发现和开发的局限性可以通过使用基于AI的工具和技术来解决。此外,机器学习让研究人员使用生物学大数据库时更加得心应手。本文简要综述了AI算法和传统化学相结合以提高药物发现的效率以及人工智能在药物发现过程中的应用研究进展,以期为我国药物发现AI的应用提供一定参考。

## 1 先导化合物的筛选

AI在筛选新的和潜在的先导化合物方面发挥着巨大的作用。目前的数据库中有大约1.06亿个化学结构,分别来自不同类型的研究,如基因组研究、临床和非临床研究、体内分析和微阵列分析研究等。根据其活性位点、结构及靶标结合能力,可利用机器学习模型,如强化模型、Logistic模型、回归模型和生成模型来筛选这些化学结构。

因可节约时间和成本,AI成为初级和次级药物筛选过程中急需的技术<sup>[21]</sup>。一些繁杂的任务,诸如细胞分类、设计新化合物、以及预测目标分子的三维结构等在AI的辅助下可以有效缩短研发时间,从而加快药物发现过程<sup>[22-23]</sup>。初级药物筛选包括通过AI技术对细胞进行分类<sup>[24]</sup>。为了对目标细胞进行分类,首先需要对ML模型进行训练,以便识别细胞及其特征<sup>[25]</sup>。次级药物筛选包括分析化合物的物理性质、生物活性和毒性。ChEMBL、PubChem和ZINC等公开数据库涵盖了数百万种化合物及其具体信息,例如结构、已知靶点等。匹配分子对(MMP)和ML可以预测生物活性,如口服暴露、体内清除、ADMET和作用方式等<sup>[26-28]</sup>。

## 2 多肽合成和小分子药物的发现

近年来,研究人员利用AI来合成新的多肽。Yan等<sup>[29]</sup>在2020年开发了基于DL的短抗菌肽鉴定

平台 Deep-AmPEP30, 并利用该平台从存在于胃肠道的光滑梭菌的基因组序列中鉴定出新的抗菌肽。此外, Kavousi 等<sup>[30]</sup>开发了用于鉴定抗菌肽的网页服务器——IAMPE, 可识别新的抗菌肽。Yi 等<sup>[31]</sup>在 2019 年设计了 ACP-DL, 这是基于 DL 的工具, 使用 LSTM 算法来发现新的抗癌肽。利用基于 AI 的工具也可以用来探索小分子药物的治疗作用。Zhavoronkov 等<sup>[32]</sup>设计了一种基于生成性强化学学习的小分子从头设计工具 GENTRL, 并用其发现了新的 DDR1 激酶抑制剂。McCloskey 等<sup>[33]</sup>将 DNA 编码的小分子文库数据与 GraphCNN 和 RF 等 ML 模型相结合发现了新型类药小分子。Xing 等<sup>[34]</sup>整合了 XGBoost、SVM 和 DNN 来寻找与类风湿性关节炎相关靶标的小分子药物。

### 3 最佳给药剂量的确定

如何确定药物疗效最佳、毒副作用最小的剂量一直是药物设计中的难题<sup>[35]</sup>。随着 AI 的出现, 许多研究人员可利用 ML 和 DL 算法的帮助来确定最佳给药剂量。例如, Shen 等<sup>[36]</sup>开发了基于 AI 的 AI-PRS 平台, 并以此确定抗逆转录病毒用于治疗艾滋病的最佳和联合用药剂量。在 10 名艾滋病患者联合使用替诺福韦、法韦伦和拉米夫定, 利用 AI-PRS 分析表明替诺福韦起始剂量可降低 33% 且不会导致病毒复发。Julkunen 等<sup>[37]</sup>使用新的 ML 驱动工具 comboFM 确定了抗癌药物克里唑替尼和硼替佐米联合用药, 并在淋巴瘤细胞系中显示出良好的疗效。Tang 等<sup>[38]</sup>使用 ML 技术, 如人工神经网络、贝叶斯加性回归树、增强回归树、多元自适应回归样条确定了免疫抑制药物他克莫司的最佳给药剂量。

### 4 类药化合物的设计和药物不良反应的预测

类药化合物的设计复杂且困难。近年来, 多种基于 AI 开发在线工具可用来分析类药化合物的释放, 以及评价所选定的具有生物活性化合物作为载体的可行性。其中最常用的是基于化学特征的药效团评价, 研究人员可利用 AI 来确定用于与某一疾病相关的特定靶标的生物活性药物。例如, Wu 等<sup>[39]</sup>利用集成 DL 和 RF 的方法设计了 WDL-RF 用于确定靶向配体的 G 蛋白偶联受体 (GPCRs) 的生物活性。还可利用 AI 在药物上市前确定其可能存在的药物不良反应。例如, Dey 等<sup>[39]</sup>使用了基于 DL 的模型来预测与药物相关的不良反应, 甚至可以识别导致这些不良反应的化学子结构。此外, Liu 等<sup>[41]</sup>通过 ML 分析, 整合了药物的化学、生物学和表型特性来预测与之相关的不良反应。Jamal 等<sup>[42]</sup>整

合了生物学、化学和表型特性, 通过 ML 分析预测与药物相关的神经系统不良反应, 并找出与针对阿尔茨海默病药物相关的 ADR。

### 5 蛋白质间相互作用的预测

蛋白质间相互作用 (PPI) 的预测对于药物发现和开发至关重要。PPI 在几乎所有的生物学过程中都发挥重要作用, 包括信号转导、细胞生长和免疫防御等。鉴于这些相互作用在内稳态和疾病反应中的关键作用, 与体内蛋白质相互作用的合成蛋白质 (例如工程化抗体) 代表了现代医学中最具变革性的治疗方法之一。

大多数合成蛋白质都是在实验平台上开发完成的, 然而开发者很难知晓这些蛋白质在哪里以及如何与靶标蛋白结合。尽管计算设计方法取得了许多进步, 但预测与靶标相互作用区域的氨基酸序列仍是结构生物学中最具挑战性的问题之一。随着研究人员对 PPI 的理解的增强, 蛋白质结构数据可用性的增加, 以及 ML 的进步, 为分析 PPI 的方法奠定了基础。比如使用贝叶斯网络 (BN) 预测 PPI。其本质是利用基因共表达、基因本体 (GO) 和其他生物学过程的相似性, 集成数据集产生精确的 PPI 网络, 展示全面的酵母相互作用组。已有研究小组使用 BN 结合酵母的数据集研究出一种新的分层模型 PCA 集成极限学习机器 (PCA-EELM), 该工具可通过仅使用蛋白质序列信息来预测蛋白质间的相互作用<sup>[43]</sup>。

### 6 虚拟筛选 (VS) 效率的提高

VS 是实验性 HTS 对应的计算方法, 即在计算机中虚拟测试化学文库中的化合物对可能与特定疾病有治疗意义的生物分子目标的活性, 其本质是从海量的化合物中选取可能的活性化合物, 在众多的化合物中只有部分是该靶标的活性化合物, 如果直接进行筛选, 虽然得到活性化合物的数量比较全, 但是得到活性化合物的概率比较低, 付出的成本也比较大。如果采用 VS 的策略, 得到活性化合物的概率大大增加。因此, VS 是药物设计和发现过程中 CADD 的重要方法之一, 也是从化合物数据库中筛选出具有治疗前景化合物的有效方法。然而, VS 耗费的成本较高且准确率低。在 VS 中引入 ML, 能够加快 VS 的速度, 并降低 VS 的假阳性率。VS 可分两种, 即基于结构的 VS (SBVS) 和基于配体的 VS (LBVS)<sup>[44-45]</sup>。其中, 分子对接是 SBVS 中应用的主要原理, 现已开发了多种基于 AI 和 ML 的评分算法, 如 NNScore、CScore、SVR-SCORE 和 ID-

SCORE<sup>[46]</sup>;也有算法用于SBVS中的分子动态模拟分析以及预测SBVS中蛋白质-配体的亲和力,如RFs、SVMs、CNNs和浅层神经网络<sup>[47]</sup>。LBVS也开发了不同的算法和工具,例如SwissSimilarity、METADOCK、HybridSim-VS、PKRank、BRUSELAS和AutoDock Bias<sup>[48-49]</sup>等。SBVS和LBVS降低了识别致病靶点的潜在治疗性化合物的复杂性,使其更为便捷精准。

## 7 QSAR建模和药物重新定位

研究化合物的结构和理化性质与生物学活性之间的关系在在药物设计和开发中非常重要。QSAR建模是利用理论计算和统计分析方法在化学结构和生物学活性之间建立定量的数学模型。QSAR模型大致分为2类,即回归模型和分类模型。目前已经开发了多种基于网络的工具和算法,如VEGA平台、QSAR-Co、FL-QSAR、Transformer-CNN和Chemception等,为QSAR建模提供了一条新的途径<sup>[50-53]</sup>。

从已获批的临床药物中有效识别新的适应症药物,可以绕过开发新药所需的多项批准前测试,该过程被称为药物重新定位。基因组学、蛋白质组学、体内和体外药理学研究中大数据集的出现为药物重新定位提供了便捷的途径。近年来,ML算法用新的系统生物学方法取代了基于化学相似性和分子对接的传统方法,而且基于AI算法和基于网络的工具的出现为该领域研究提供了平台,如DrugNet、DRIMC、DPDR-CPI、PHARMGKB和DRRS等<sup>[54-58]</sup>。Hooshmand等<sup>[59]</sup>基于神经网络进行药物重新定位,确定了16种潜在的抗新型冠状病毒的可再利用药物,并基于多模型DL方法为新型冠状病毒肺炎确定了12个具有前景的药物靶标。

## 8 理化性质和药物靶标亲和力的预测

溶解度、分配系数、电离度、渗透系数等理化性质可能会影响化合物的药代动力学特性和药物-靶标结合效率。因此,在设计药物分子时,必须考虑化合物的理化性质。目前已经开发了多种基于AI的工具来预测化合物的上述理化性质,例如分子指纹分析、SMILES格式、库仑矩阵和势能测量,这些工具均用于DNN训练阶段<sup>[60, 61]</sup>。

预测化学分子与治疗靶标的结合亲和力是药物发现和开发的重要环节,而AI算法的最新进展则加速了该过程。研究人员已利用药物及其靶点的相似性特征开发了一些基于网络的工具,如ChemMapper和相似集成法(SEA)。此外,还构建

了基于ML和DL的药物靶标亲和力识别模型,如KronRLS、SimBoost、DeepDTA和Padme等<sup>[62]</sup>。

## 9 化合物的结合预测和体内安全性分析

AI可以在合成前预测药物分子与靶标结合及未结合时的效应,以及进行体内安全性分析。近年来DL算法帮助研究人员开发出能够分析化合物分子表征且适用于预测化合物毒性的DL方法,例如DL在抗菌药物发现中的应用可以帮助选择新的强效先导化合物,其具有理想的药理学和毒性特征,以供进一步优化。随着计算能力和算法的不断改进,可供使用的数据量的也不断增加<sup>[46-48, 63]</sup>。DL算法在毒性预测方面的潜力取决于数据集的质量和数量,因此尚需进行更深入的研究以使基于AI的算法对毒性预测更加可靠。

## 10 多靶点配体药物分子的设计

AL和ML算法在药物发现和开发中的重要成果之一是预测和估计疾病网络、药物间相互作用或药物-靶标关系的整体拓扑和动力学。DisGeNET、STRITCH等数据库分别被用于确定基因-疾病关联、药物-靶标关联和分子通路。例如,Gu等<sup>[64]</sup>在2020年使用相似集成法确定了197种最常用中药的靶点,然后使用DisGeNET数据库将这些靶标与不同疾病联系起来,从而将中药与可治疗的疾病联系起来。多重药理学最适合于对复杂性疾病,如癌症、神经退行性疾病、糖尿病和心力衰竭等治疗药物的设计。基于ML的方法具有分析关联推断分子网络的潜力,极大地增加了发现多靶点配体药物分子的概率。此外,ML模型还有助于识别具有不同结合位点的多靶点配体药物分子。

## 11 临床试验的设计

临床试验是新药研发过程中周期最长、成本最高的环节。目前,全球各国药物临床试验阶段的成功率均较低,严重影响了药物研发并浪费了大量时间和费用。在临床试验环节,可以利用AI技术辅助临床试验设计、患者招募和临床试验数据处理。例如IBM Watson开发的临床试验配对系统<sup>[65]</sup>通过利用患者的医疗记录和之前的大量临床试验数据来创建详细档案。AI模型还可以通过分析毒性、副作用和其他相关参数来提高药物临床试验的成功率,从而降低临床试验的费用。在不远的将来,AI技术还将促进临床试验数据的良好管理,进而实现个性化医疗的目标。

## 12 结语

AI正在成为解决医学、生命科学和工程领域复

杂问题的强有力手段。综上所述, AI可参与药物开发过程的各个阶段, 为新药开发提供了“发现-上市”的一站式服务, 如先导化合物的筛选、多肽合成及小分子药物的发现、最佳给药剂量的确定、类药化合物的设计和不良反应的预测、蛋白质间相互作用的分析、虚拟筛选效率的提高、QSAR建模和药物重新定位、理化性质和药物靶标亲和力的预测、化合物的结合预测和体内安全性分析、多靶点配体药物分子的设计至辅助临床试验设计。ML和DL方法的最新进展为药物发现和开发带来了“降本增效”的好机会, 而AI算法的进步, 特别是DL方法的进步, 以及架构硬件的改进和大数据的更易获得, 都表明AI的第三波浪潮正在袭来, 许多制药公司也相继与AI公司开展合作。当前, AI在药物发现领域已被成功用于靶标识别、先导化合物优化、ADMET预测、临床试验设计等方面。2020年12月, Insilico Medicine公司的小分子抑制剂向美国FDA提出临床研究申请, 且已于2022年底完成I期临床, 并在2023年获得美国FDA授予的孤儿药资格。该药物是有史以来第一次通过基于AI工具发现并获批的新靶标小分子抑制剂。

虽然目前AI在药物发现领域已经有一些成功案例, 但在高质量数据的获取方面仍存在2个主要挑战。首先, 标记不能是二元的, 因为药物在生物系统中的作用是复杂的; 其次, 虽然数据库拥有海量信息, 但药物发现中可用的高质量数据并不多。因此, 迫切需要建立1个不仅能提供海量数据, 而且还能提供高质量数据的平台。在制药行业, 开放数据共享并不常见, 皮斯托亚联盟(Pistoia alliance)鼓励许多公司与他人共享数据, 并准备建立统一的数据格式, 但是目前尚需解决技术上的难题。另一方面, 由于人类疾病的复杂性和人体的特异性, 将AI工具用于药物研发过程中也存在一些不可避免的困难。因此, AI技术和多种算法需要投入更多资金和研究工作, 并需要与传统基础类学科和临床医学各学科有机地结合。但不可否认的是, 在不久的将来, AI将会给药物发现和开发过程带来革命性的变化。希望本文简要综述的人工智能在药物发现过程中的应用研究进展, 能够为中国药物发现的AI应用提供一定参考。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

[1] Lipinski C F, Maltarollo V G, Oliveira P R, et al.

Advances and perspectives in applying deep learning for drug design and discovery [J]. *Front Robot AI*, 2019, doi: 10.3389/frobt.2019.00108.

- [2] Hassanzadeh P, Atyabi F, Dinarvand R. The significance of artificial intelligence in drug delivery system design [J]. *Adv Drug Deliv Rev*, 2019, 151-152: 169-190.
- [3] Zhong F, Xing J, Li X, et al. Artificial intelligence in drug design [J]. *Sci China Life Sci*, 2018, 61(10): 1191-1204.
- [4] Badillo S, Banfai B, Birzele F, et al. An introduction to machine learning [J]. *Clin Pharmacol Ther*, 2020, 107(4): 871-885.
- [5] Duch W, Swaminathan K, Meller J, et al. Artificial intelligence approaches for rational drug design and discovery [J]. *Curr Pharm Des*, 2007, 13(14): 1497-1508.
- [6] Brown N, Ertl P, Lewis R. et al. Artificial intelligence in chemistry and drug design [J]. *J Comput Aided Mol Des*, 2020, 34(7): 709-715.
- [7] Russell S, Norvig P. *Artificial Intelligence: A Modern Approach* [M]. 3rd Ed. Englewood Cliffs: Prentice-Hall, Inc., 2010.
- [8] Ron E, Michael D, Lash A E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository [J]. *Nucleic Acids Res*, 2002, 30(1): 207-210.
- [9] Wang Z, Jensen M A, Zenklusen J C. A practical guide to the cancer genome atlas (TCGA) [J]. *Methods Mol Biol*, 2016, 1418: 111-141.
- [10] van Zendoorn I D G P, Szuhai K, de Bruijn B I H, et al. Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas [J]. *PLoS Comput Biol*, 2019, 15(2): e1006826.
- [11] Bento A P, Gaulton A, Hersey A, et al. The ChEMBL bioactivity database: An update [J]. *Nucleic Acids Res*, 2014, 42(D1): D1083-D1090.
- [12] Mendez D, Gaulton A, Bento A P, et al. ChEMBL: towards direct deposition of bioassay data [J]. *Nucleic Acids Res*, 2019, 47(D1): D930-D940.
- [13] Wishart D S, Feunang Y D, Guo A C, et al. DrugBank 5.0: A major update to the DrugBank database for 2018 [J]. *Nucleic Acids Res*, 2017, 46(D1): D1074-D1082.
- [14] Keenan A B, Jenkins S L, Jagodnik K M, et al. The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations [J]. *Cell Syst*, 2018, 6(1): 13-24.
- [15] Fan Y, Zhang Y, Hua Y, et al. Investigation of machine intelligence in compound cell activity classification [J]. *Mol Pharm*, 2019, 16(11): 4472-4484.
- [16] Zhang D H, Wu K L, Zhang X, et al. In silico screening of Chinese herbal medicines with the potential to directly

- inhibit 2019 novel coronavirus [J]. *J Integr Med*, 2020, 18(2): 152-158.
- [17] Álvarez-Machancoses Ó, Fernández-Martínez J L. Using artificial intelligence methods to speed up drug discovery [J]. *Expert Opin Drug Discov*, 2019, 14(8): 769-777.
- [18] Fleming N. How artificial intelligence is changing drug discovery [J]. *Nature*, 2018, 557(7707): S55-S57.
- [19] Segler M H S, Kogej T, Tyrchan C, et al. Generating focused molecule libraries for drug discovery with recurrent neural networks [J]. *ACS Cent Sci*, 2018, 4(1): 120-131.
- [20] 梁礼, 邓成龙, 张艳敏, 等. 人工智能在药物发现中的应用与挑战 [J]. *药学进展*, 2020, 44(1): 18-27.  
Liang L, Deng C L, Zhang Y M, et al. Application and challenges of artificial intelligence in drug discovery [J]. *Progr Pharm Sci*, 2020, 44(1): 18-27.
- [21] Szklarczyk D, Santos A, von Mering C, et al. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data [J]. *Nucleic Acids Res*, 2016, 44: D380-D384.
- [22] Zhavoronkov A, Vanhaelen Q, Oprea T I. Will artificial intelligence for drug discovery impact clinical pharmacology? [J]. *Clin Pharmacol Ther*, 2020, 107(4): 780-785.
- [23] Watson O P, Cortes-Ciriano I, Taylor A R, et al. A decision-theoretic approach to the evaluation of machine learning algorithms in computational drug discovery [J]. *Bioinformatics*, 2019, 35(22): 4656-4663.
- [24] 陈长亮. 基于机器学习的药物筛选辅助方法研究 [D]. 大连: 大连理工大学, 2022.  
Chen C L. Research on drug screening auxiliary method based on machine learning [D]. Dalian: Dalian University of Technology, 2022.
- [25] Tripathy R K, Mahanta S, Paul S. Artificial intelligence-based classification of breast cancer using cellular images [J]. *RSC Adv*, 2014, 4(18): 9349-9355.
- [26] Chan H C S, Shan H, Dahoun T, et al. Advancing drug discovery via artificial intelligence [J]. *Trends Pharmacol Sci*, 2019, 40(8): 592-604.
- [27] Carpenter K A, Huang X. Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: A review [J]. *Curr Pharm Des*, 2018, 24(28): 3347-3358.
- [28] Schyman P, Liu R, Desai V, et al. vNN web server for ADMET predictions [J]. *Front Pharmacol*, 2017, 8: 889.
- [29] Yan J, Bhadra P, Li A, et al. Deep-AmPEP30: improve short antimicrobial peptides prediction with deep learning [J]. *Mol Ther Nucleic Acids*, 2020, 20: 882-894.
- [30] Kavousi K, Bagheri M, Behrouzi S, et al. IAMPE: NMR-Assisted computational prediction of antimicrobial peptides [J]. *J Chem Inf Model*, 2020, 60(10): 4691-4701.
- [31] Yi H C, You Z H, Zhou X, et al. ACP-DL: A deep learning long short-term memory model to predict anticancer peptides using high-efficiency feature representation [J]. *Mol Ther Nucleic Acids*, 2019, 17: 1-9.
- [32] Zhavoronkov A, Ivanenkov Y A, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors [J]. *Nat Biotechnol*, 2019, 37(9): 1038-1040.
- [33] McCloskey K, Sigel E A, Kearnes S, et al. Machine learning on DNA-encoded libraries: A new paradigm for hit finding [J]. *J Med Chem*, 2020, 63(16): 8857-8866.
- [34] Xing G, Liang L, Deng C, et al. Activity prediction of small molecule inhibitors for antirheumatoid arthritis targets based on artificial intelligence [J]. *ACS Comb Sci*, 2020, 22(12): 873-886.
- [35] Dimmitt S, Stampfer H, Martin J H. When less is more - efficacy with less toxicity at the ED<sub>50</sub> [J]. *Br J Clin Pharmacol*, 2017, 83(7): 1365-1368.
- [36] Shen Y, Liu T, Chen J, et al. Harnessing artificial intelligence to optimize long-term maintenance dosing for antiretroviral-naive adults with HIV-1 infection [J]. *Adv Ther*, 2020, doi:10.1002/adtp.201900114.
- [37] Julkunen H, Cichonska A, Gautam P, et al. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects [J]. *Nat Commun*, 2020, doi: 10.1038/s41467-020-19950-z.
- [38] Tang J, Liu R, Zhang Y L, et al. Corrigendum: application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients [J]. *Sci Rep*, 2018, doi: 10.1038/srep46936.
- [39] Wu Z, Chen L. Similarity-based method with multiple-feature sampling for predicting drug side effects [J]. *Comput Math Methods Med*, 2022, doi: 10.1155/2022/9547317.
- [40] Liu X, Gao Y, Peng J, et al. TarPred: A web application for predicting therapeutic and side effect targets of chemical compounds [J]. *Bioinformatics*, 2015, 31(12): 2049-2051.
- [41] Liu M, Wu Y, Chen Y, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs [J]. *J Am Med Inform Assoc*, 2012, 19(e1): e28-e35.
- [42] Jamal S, Goyal S, Shanker A, et al. Predicting neurological adverse drug reactions based on biological, chemical and phenotypic properties of drugs using machine learning models [J]. *Sci Rep*, 2017, doi: 10.1038/s41598-017-00908-z.
- [43] You Z H, Lei Y K, Zhu L, et al. Prediction of protein-

- protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis [J]. *BMC Bioinformatics*, 2013, doi: 10.1186/1471-2105-14-S8-S10.
- [44] Gonczarek A, Tomczak J M, Zaręba S, et al. Interaction prediction in structure-based virtual screening using deep learning [J]. *Comput Biol Med*, 2018, 100: 253-258.
- [45] Goh G B, Hodas N O, Vishnu A. Deep learning for computational chemistry [J]. *J Comput Chem*, 2017, 38 (16): 1291-1307.
- [46] 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.
- [47] Serafim M S M, Kronenberger T, Oliveira P R, et al. The application of machine learning techniques to innovative antibacterial discovery and development [J]. *Expert Opin Drug Discov*, 2020, 15(10): 1165-1180.
- [48] Melville J L, Burke E K, Hirst J D, et al. Machine Learning in Virtual Screening [J]. *Comb Chem High Throughput Screen*, 2009, 12(4): 332-343.
- [49] Singh N, Chaput L, Villoutreix B O. Virtual screening web servers: Designing chemical probes and drug candidates in the cyberspace [J]. *Brief Bioinform*, 2021, 22(2): 1790-1818.
- [50] Simm J, Klambauer G, Arany A, et al. Repurposing high-throughput image assays enables biological activity prediction for drug discovery [J]. *Cell Chem Biol*, 2018, 25 (5): 611-618.
- [51] Czyzewski A, Krawiec F, Brzezinski D, et al. Detecting anomalies in X-ray diffraction images using convolutional neural networks [J]. *Expert Syst Appl*, 2021, doi: 10.1016/j.eswa.2021.114740.
- [52] Preuer K, Lewis R P L, Hochreiter S, et al. DeepSynergy: Predicting anti-cancer drug synergy with deep learning [J]. *Bioinformatics*, 2018, 34(9): 1538-1546.
- [53] Luechtefeld T, Marsh D, Rowlands C, et al. Machine learning of toxicological big data enables read-across structure activity relationships (RASAR) outperforming animal test reproducibility [J]. *Toxicol Sci*, 2018, 165(1): 198-212.
- [54] Zhang W, Xu H, Li X, et al. DRIMC: An improved drug repositioning approach using Bayesian inductive matrix completion [J]. *Bioinformatics*, 2020, 36(9): 2839-2847.
- [55] Luo H, Zhang P, Cao X H, et al. DPDR-CPI, a server that predicts drug positioning and drug repositioning via chemical-protein interactome [J]. *Sci Rep*, 2016, doi: 10.1038/srep35996.
- [56] Zhu Q, Tao C, Shen F, et al. Exploring the pharmacogenomics knowledge base (PharmGKB) for repositioning breast cancer drugs by leveraging Web ontology language (OWL) and cheminformatics approaches [A]// 18th Pacific Symposium on Biocomputing [C]. Big Island of Hawaii: International Society for Computational Biology, 2014.
- [57] Gallo K, Goede A, Eckert A, et al. PROMISCUOUS 2.0: A resource for drug-repositioning [J]. *Nucleic Acids Res*, 2021, 49(1): D1373-D1380.
- [58] Luo H, Li M, Wang S, et al. Computational drug repositioning using low-rank matrix approximation and randomized algorithms [J]. *Bioinformatics*, 2018, 34(11): 1904-1912.
- [59] Hooshmand S A, Zarei Ghobadi M, Hooshmand S E, et al. A multimodal deep learning-based drug repurposing approach for treatment of COVID-19 [J]. *Mol Divers*, 2021, 25(3): 1717-1730.
- [60] Schneider P, Walters W P, Plowright A T, et al. Rethinking drug design in the artificial intelligence era [J]. *Nat Rev Drug Discov*, 2020, 19(5): 353-364.
- [61] Chen H, Engkvist O, Wang Y, et al. The rise of deep learning in drug discovery [J]. *Drug Discov Today*, 2018, 23(6): 1241-1250.
- [62] Özçelik R, Öztürk H, Özgür A, et al. ChemBoost: A chemical language based approach for protein - ligand binding affinity prediction [J]. *Mol Inform*, 2021, doi: 10.1002/minf.202000212.
- [63] Nayariseri A, Khandelwal R, Tanwar P, et al. Artificial intelligence, big data and machine learning approaches in precision medicine & drug discovery [J]. *Curr Drug Targets*, 2021, 22(6): 631-655.
- [64] Gu S, Lai L H. Associating 197 Chinese herbal medicine with drug targets and diseases using the similarity ensemble approach [J]. *Acta Pharmacol Sin*, 2020, 41(3): 432-438.
- [65] Fogel D B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review [J]. *Contemp Clin Trials Commun*, 2018, 11: 156-164.

[责任编辑 李红珠]