

• 专家论坛 •

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

刘润哲, 宋俊科, 刘艾林, 杜冠华*

(中国医学科学院、北京协和医学院药物研究所, 国家药物筛选中心, 北京市药物靶点研究和
新药筛选重点实验室, 北京 100050)

摘要: 人工智能技术在药物筛选中的应用日益广泛。本文介绍了人工智能技术的特点, 着重从机器学习尤其是深度学习角度, 按照基于配体和受体结构两个方面, 总结了人工智能技术在药物筛选中的应用和进展, 以及如何应用人工智能从这两个方面进行药物设计。本文也讨论了人工智能技术在药物虚拟筛选领域的主要局限性和挑战, 对其发展前景作以展望。

关键词: 人工智能; 虚拟筛选; 计算机辅助药物设计; 药理学

中图分类号: R966 文献标识码: A 文章编号: 0513-4870(2021)08-2136-10

Progress on the application of artificial intelligence technology in ligand-based and receptor structure-based drug screening

LIU Run-zhe, SONG Jun-ke, LIU Ai-lin, DU Guan-hua*

(National Center for Pharmaceutical Screening, Beijing Key Lab of Drug Target Identification and Drug Screening, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China)

Abstract: Artificial intelligence technology is being widely applied in drug screening. This paper introduces the characteristics of artificial intelligence, and summarizes the application and progress of artificial intelligence technology especially deep learning in drug screening, from ligand-based and receptor structure-based aspects. This paper also introduces how to apply artificial intelligence to drug design from these two aspects. Finally, we discuss the main limitations, challenges, and prospects of artificial intelligence technology in the field of drug screening.

Key words: artificial intelligence; virtual screening; computer aided drug design; pharmacology

近年来, 随着药物研发领域的数字化程度不断提高, 人们应用人工智能 (artificial intelligence, AI) 技术来应对药物发现中的挑战, 以获得启发, 增加药物研发的效率。目前, 人工智能技术在药物发现、临床前研究和临床研究等领域均有应用^[1]。药物筛选处于药物研发的早期阶段^[2,3], 其目的是从大量的化合物中寻找具有特定药理活性的先导化合物。在虚拟筛选中应用人

工智能技术, 可以根据化合物的结构进行药理学性质预测, 为后续的实验研究提供方向, 是新药发现的重要辅助手段^[4]。

AI泛指模仿人类思维和认知功能的机器及相关技术^[5]。不过, 跨学科语境下的AI概念有时特指深度学习, 它其实是AI范畴中机器学习领域下的一种技术。近年来, 深度学习在棋类游戏、人脸识别、自动驾驶和疾病诊断等方面不断取得突破, 使得深度学习成为进展最快、最具代表性的AI技术。

本文探讨的AI技术以深度学习为主, 也涵盖其他的机器学习技术。机器学习一般是使用特定的算法, 通过拟合真实数据, 让计算机构建模型, 形成对人类经

收稿日期: 2021-01-11; 修回日期: 2021-03-08.

基金项目: 国家重大新药创制科技重大专项 (2018ZX09711001-012); 中国医学科学院医学与健康科技创新工程 (2020-I2M-1-003).

*通讯作者 Tel: 86-10-63165184, E-mail: dugh@imm.ac.cn

DOI: 10.16438/j.0513-4870.2021-0052

验的模仿。机器学习 (machine learning, ML) 一词最初由 Arthur Samuel 在 1959 年创造^[6], 它旨在构建能够随经验而自动改进的计算机程序^[7,8]。从 20 世纪 80 年代起, 机器学习的理论和模型取得了快速的发展, 包括人工神经网络^[9]、贝叶斯网络^[10]、支持向量机^[11]和决策树^[12]等。具有多层结构的人工神经网络被称为深度学习, 其特点是包含一组相互连接的人工神经元, 每个神经元可以接受其他神经元的输入, 通过函数计算得到一个输出值并可传递给其他神经元; 神经元之间连接的强度由“权重”表示, 通过改变权重, 神经网络可以近似拟合任何函数^[7,13]。机器学习经过几十年的发展, 技术方法都有了很大提高, 新算法不断出现, 成为辅助科学研究的重要工具。在实际应用中, 以机器学习为主的 AI 技术适用于解决已经存在大量数据、但与之相关的理论和模型仍未知的问题^[1]。

在药物筛选领域, 应用 AI 技术能够较好地模拟药理学中常见的非线性关系, 在解决有大量数据的复杂问题时具有重要的应用。近年来, AI 开源库的发展, 如 Scikit-learn^[14]、TensorFlow^[15]和 PyTorch 等, 能够让人们更方便地编写 AI 程序。在药物研发中, 以 DeepChem^[16]、AMPL^[17]和 cando.py^[18]为代表的开源库则在上述库的基础上, 结合药物研发的需求, 进一步优化 AI 的技术流程, 使 AI 技术更加灵活易用, 促进了药理学和 AI 的交叉领域的发展。

目前, 基于计算机的药物筛选主要有两种思路: 第一种思路是分析已知的活性分子的特征, 通过对比待测化合物与已知活性化合物的相似性, 预测其药理学性质, 也称基于配体的方法^[19,20]; 第二种思路是模拟化合物与蛋白质靶点之间在三维空间上的相互作用, 进而预测化合物对靶蛋白的作用方式, 也称基于受体结构的方法。两种思路各有侧重, 互为补充, 成为目前虚拟筛选的主要理论体系。实际上, 药物发现的途径还有很多, 应用计算机技术和 AI 技术进行药物发现, 还有很多需要探讨的问题。本文仅从基于配体和受体结构这两个方面探讨 AI 在药物筛选中的应用和进展。

1 AI 在基于配体的虚拟筛选中的应用

近年来, AI 在基于配体的虚拟筛选中取得了较多进展。基于配体的 AI 筛选的工作流程一般是: 从数据库或文献中收集已知的活性化合物的信息作为训练数据; 根据分子结构选择合适的表示方法; 建立生物活性和分子表示之间的关系模型; 应用模型预测待测化合物的生物特性^[21]。基于配体的 AI 方法不仅适用于预测药物和靶蛋白的相互作用, 也可用于预测化合物的理化性质、药代动力学性质和对基因表达谱的影响等。

1.1 训练数据和分子表示 数据获取是 AI 筛选流程的第一步。AI 方法的预测能力取决于训练数据质量和数量^[1], 因此, 获取丰富且可靠的数据是使用 AI 方法的重要前提。在药物筛选领域, 训练数据一般是开放的数据库中采集的, 其中, ChEMBL^[22,23]和 PubChem Bioassay^[24,25]等数据库中收录了大量的化合物药效数据, 常用作药物虚拟筛选的训练数据的来源。表 1 展示了药物筛选中常用的数据库^[22-36]。

分子表示即把分子转化成数据, 让计算机进行处理, 一般要尽可能完整、有效地表达化合物的信息。常用的分子表示方法包括简化分子线性输入规范 (simplified molecular input line entry specification, SMILES)^[37]、分子描述符^[38]、分子指纹^[39]和分子图 (graph) 等, 都可以取得良好的效果。如同 AI 技术本身, 分子表示方法也处在不断的发展和改进之中^[40]。目前, SMILES 是 AI 中最流行的分子表示方法, 这种分子表示方法有时会进一步地经由 RDKit 包等工具形成 AI 模型所需的其他分子特征, 如分子指纹或描述符等。在虚拟筛选的工作中, 可以使用 Open Babel^[41]等软件, 将其他形式的分子格式转化为 SMILES 的形式, 供 AI 程序进行处理。

1.2 预测化合物的生物学效应 虚拟筛选的主要目的之一是采用计算机预测化合物的生物学效应。而在虚拟筛选中, 评价生物学效应的基础是药物与靶点的相互作用。在基于配体的方法中, 一般假定相似的化合物能够与相似的靶点作用, 并以此构建各种 AI 模型。例如, 回归是一种基础的 AI 算法, 基于正则最小二乘回归的 KronRLS 模型可以用来预测药靶作用的连续值, 不受限于有活性和无活性的简单分类^[42]。SimBoost 则进一步地利用回归树预测药靶相互作用, 并考虑基于特征和基于相似性的相互作用^[43]。应用深度学习技术的例子包括基于卷积神经网络 (convolutional neural network, CNN) 的 WideDTA、DeepDTA^[44]、DEEPscreen^[45], 基于图卷积神经网络的 PADME, 基于递归神经网络的 DeepAffinity^[46]等。这些方法在预测药物-靶点相互作用时, 一般以 SMILES 表示化合物, 以氨基酸序列等形式输入靶点蛋白的信息。如 Majumdar 等^[47]从药靶相互作用的角度, 以 S 糖蛋白为靶点, 寻找潜在的抗 SARS-CoV-2 新冠病毒的药物。在这里, 作者以激酶抑制数据集作为训练数据, 靶点和配体分别以氨基酸序列和 SMILES 序列输入; 靶点信息进一步地经由算法转化为 8 420 维的向量, 而配体信息则进一步地转化成 1 024 维的向量, 二者合并为一个 9 444 维的向量; 这些数据用来训练一个一维的卷积神经网络模型, 该模型被用于预测化合物列表对 S-糖蛋白的抑制作用, 预测了多个潜在 S-糖蛋白抑制剂。

Table 1 Frequently-used databases in drug discovery

Database	Url and included data	Ref.
BingingDB	http://www.bindingdb.org Affinity information of small molecules and proteins	[26]
CancerDR	http://crdd.osdd.net/raghava/cancerdr Anticancer drugs and their effectiveness against cancer cell lines	[27]
CancerResource	http://bioinformatics.charite.de/care Drug-target relationships related to cancer	[28]
CellMiner	http://discover.nci.nih.gov/cellminer Molecular and pharmacological data for the NCI-60 cancerous cell lines	[29]
ChemBank	http://chembank.broadinstitute.org Molecular descriptor and compound information	[30]
ChEMBL	https://www.ebi.ac.uk/chembl Chemical, biological activity, and genomic data of compounds	[22,23]
CMap	https://portals.broadinstitute.org/cmap Gene expression data of human cell lines induced by compounds	[31]
DrugBank	http://www.drugbank.ca Effects, targets, and other information of drugs	[32]
KEGG pathway	https://www.genome.jp/kegg/pathway.html Pathways of pathology, physiology, and drug action	[33]
Open Targets	https://www.targetvalidation.org Association of gene and disease	[34]
Pathway Commons	http://www.pathwaycommons.org Multi-species biological pathway data	[35]
PubChem bioassay	http://pubchem.ncbi.nlm.nih.gov Biological activity data of compounds from high-throughput screening	[24,25]
TTD	http://db.idrblab.net/ttd Target, disease, and pathway information	[36]

值得一提的是,基于神经网络的方法,其预测效果并不必然优于其他原理更为简单的AI模型。一般认为,神经网络的方法在训练数据的数量和质量充足的情况下可能具有最好的预测效果,而在实际应用中,模型的准确率与待解决问题的特性、训练数据以及采用的方法或设计都有关联。

除了化合物与靶点的相互作用的角度外,也有研究从化合物诱导基因表达谱的角度,以及化合物对细胞、组织或个体的表型作用角度来评价化合物的生物学效应。在化合物诱导的基因表达的方面,一般是通过对CMap^[31]平台及其改进版本L1000^[48]获取已知的药物在细胞中诱导基因表达的信息作为训练数据,通过对比化合物与这些药物在多个方面的相似性,预测化合物的作用,常用于药物再利用的研究^[2,49-54]。Kim等^[55]使用CMap来发现抗衰老药物,首先从GEO数据库获取并分析了30岁以下(青年组)和50岁以上(老年组)急性髓细胞白血病患者基因表达谱的差异,将差异基因视为抗衰老的关键基因;同时获取了CMap中,不同化合物诱导HL-60细胞(一种白血病细胞株)的基因表达谱的数据;利用TensorFlow编写神经网络模型,以构建化合物结构与上述抗衰老关键基因表达的关联,预测化合物的抗衰老作用;发现了曲古抑菌素A和伏立诺他等药物及药物组合具有潜在的抗

衰老作用。Carrella等^[56]开发了MANTRA平台,它可以根据用户上传的化合物给药前后细胞的基因表达谱数据,利用CMap数据集,预测该化合物的药物作用网络,并对具有相近作用机制的化合物进行聚类。

在表型作用方面,除了从数据库中获取训练数据外,还可以从患者的电子健康记录中获取训练数据^[57,58],或以大型药厂公开的数据集作为训练数据,或者结合深度学习领域的计算机视觉技术,将高通量成像数据作为表型数据的来源进行药物发现^[59]。

1.3 预测化合物的吸收、分布、代谢、排泄性质和耐受毒性(ADMET)性质除了药效性质外,理想的药物应当具备优良的ADMET^[60,61],ADMET的预测也是指导先导化合物优化工作的关键^[62]。常用于ADMET预测的AI方法包括k近邻、支持向量机和随机森林等^[62-66],它们与神经网络相比具有较低复杂度和较高可解释性^[67]。而基于神经网络的AI模型^[68-70]则追求更高的预测准确性。目前,一些基于AI的ADMET预测工具包括预测CYP450反应物的CypReact^[71],预测PK性质的FAF-Drugs4^[72],预测化合物代谢稳定性的MetStabOn^[73],预测理化性质和药代性质的SwissADME^[74]、HitDexter^[75],预测ADMET性质的vNN-ADMET^[76]、ADMETlab^[77],预测化合物毒性的DeepTox^[78]、TOP^[79],预测hERG阻断能力的hERG-Att^[80]等。

与基于受体结构的方法相比, AI在基于配体的方法中的应用较为广泛, 因为化合物的生物活性数据往往比蛋白质的晶体结构更容易获得^[81]。此外, 基于配体的虚拟筛选方法一般计算量更小, 速度更快, 适合大规模、高通量的虚拟筛选。然而, 必须注意的是, 基于配体的筛选存在“活性悬崖问题”, 即具有相似结构的化合物常常表现出明显不同的活性^[82], 这对基于配体的方法提出了挑战。因此, 通过计算手段预测小分子与靶蛋白的相互作用时, 可以将不同的方法互为补充, 以增加预测结果的可靠性。

2 人工智能在基于受体结构的虚拟筛选中的应用

基于受体结构的虚拟筛选方法, 一般是通过特定算法模拟化合物和蛋白质的相互作用, 并且依赖已知的蛋白质三维结构。其中最常用的方法是分子对接方法, 可用于预测小分子与蛋白质的结合方式和结合强度^[83]。

2.1 改善分子对接算法

AI在基于受体结构的虚拟筛选中, 主要用于改善分子对接的评分函数。目前, 尽管常用的分子对接程序能够还原接近共结晶结构的分子对接形式, 但在虚拟筛选时, 分子对接技术的准确性有限, 预测的高结合力配体常常不能在生物学实验中产生活性。这一问题的原因之一来自评分函数^[84], 如许多评分函数没有考虑溶剂化和熵效应^[85], 对蛋白质柔韧性考虑欠佳^[84], 忽略结合的停留时间等^[86]。AI可以改良分子对接的评分函数, 因为随着训练样本的增加, 在理论上, AI的准确性能够不断提升, 并可以超越传统的评分函数^[87,88]。常用的AI模型主要是随机森林模型^[87,89-91]和卷积神经网络。应用3D卷积神经网络的研究一般是以实验测定的化合物-靶点相互作用数据或DUD-E数据集进行训练。Torng等^[92]使用无监督学习算法来寻找蛋白质的活性口袋的最佳表示方法, 并使用图卷积神经网络, 分别从蛋白质口袋和2D配体图中提取特征, 捕捉到药-靶相互作用。其他比较有代表性的研究成果包括早期的AtomNet, 以及已开发为网页版对接工具的BindScope^[93]、K_{DEEP}^[94]和DeltaDelta^[95]等。这些评分工具的准确率优于以Surflex-Dock^[96]和Dock^[97]为代表的传统评分算法。然而, 目前仍缺乏大规模的基于活性实验的研究来比较这些对接算法的准确性。在计算速度方面, 基于AI的分子对接算法一般比传统的对接算法速度更快, 因此在高通量的分子对接中具有一定优势。

2.2 蛋白质结构的预测

明确靶蛋白的三维结构和活性位点是基于受体结构的虚拟筛选的重要前提。已测定的蛋白质的结构可以从PDB (<https://www.rcsb.org/>) 等数据库中获取, 然而通过实验方法来测定蛋白质结构, 其难度较大。传统的计算方法, 即同源建模

和物理建模, 也有各自的局限性^[16,84]。因此, 研究人员尝试使用AI从一级序列预测蛋白质三维结构。这里, 比较有代表性的研究是基于卷积神经网络AI模型的AlphaFold^[98], 以及AlQuraishi^[99]开发的基于循环几何神经网络的模型, 均取得了超过传统计算方法的准确率和计算速度。此外, 一些研究使用AI的方法预测靶点的可成药性, 或从蛋白质的三维结构寻找活性口袋^[100-108]。这些领域进展有望为基于受体结构的虚拟筛选方法铺设更广泛的道路。

3 人工智能生成先导化合物

虚拟筛选是从一个已知的、有限的化合物库中寻找具有所需性质的化合物。那么当AI模型自身能够生成化合物的结构, 去创造“化合物库”时, AI方法就能够生成潜在的先导化合物。基于配体的思路是将已知的活性分子作为训练集, 使用AI工具总结其特征, 并生成相似的新分子。常用的模型包括递归神经网络 (recurrent neural network, RNN)、生成式对抗网络 (generative adversarial network, GAN) 和变分自编码器 (variational auto-encoder, VAE), 以及与强化学习技术相结合的方法^[109-111]。RNN和GAN主要用于产生具有期望生物活性的分子^[112-117]。而VAE模型不仅可以生成化合物分子, 还可以针对特定的期望性质优化化合物的结构^[118-120]。Merk等^[121]使用RNN设计了维甲酸X受体 (retinoid X receptor, RXR) 和过氧化物酶体增植物激活受体 (peroxisome proliferator activated receptor, PPAR) 的激动剂, 在细胞实验中可以在纳摩尔水平表现出生物活性。Zhavoronkov等^[122]应用多种神经网络, 结合强化学习技术开发了名为GENTRL的模型, 生成DDR1激酶的抑制剂, 其中部分化合物在体内和动物水平均体现了令人满意的药效, 并在激酶谱的测试中表现出良好的选择性。Yang等^[123]利用长短期记忆人工神经网络 (一种RNN的变体) 生成化合物结构, 并进行优化, 开发出了有效浓度在纳摩尔级的p300/CBP靶点抑制剂。Blaschke等^[124]则开发了一个称为REINVENT的开源的Python应用, 帮助科研人员使用AI生成潜在的药物分子。上述的例子是采用基于配体的思想构建生成模型, 生成的化合物往往与已知的活性化合物结构相似, 这在一定程度上限制了生成分子的新颖性。

相应地, 一些研究尝试其他的思路来生成高创新性的活性分子。Méndez-Lucio等^[117]从系统生物学的角度, 以化合物的结构和化合物诱导基因表达谱的数据训练GAN模型, 该模型可根据所需的基因表达谱来生成新的活性分子。Skalic等^[125]则采用基于受体结构的药物设计思路, 使用GAN来生成与蛋白质口袋互补

的配体结构。目前, AI生成的活性分子与真正的药物还存在距离, 这些模型对化合物的稳定性、合成可行性、药代动力学性质和毒性等的考虑仍然不足。这些工具的主要应用是设计一个命中率更高的虚拟化合物库, 以提高药物筛选的效率。

4 AI在药物筛选中的主要挑战

AI的药物筛选技术仍有待发展。目前, AI在药物研发中尚处于起步阶段, 其对药物发现的价值还需要时间和成果来证明。在方法开发方面, 需要将药物发现中的问题进行合理抽象, 把它对应到AI已经能处理的经典问题上去; 在工具使用方面, 如何将AI的优势最大化利用, 也值得药物研发人员的思考。

药学数据的特殊性对AI技术提出更高要求。AI在药物筛选中的一大挑战来自训练数据的数量和质量, 它对AI的效果具有决定性的影响^[1]。在药物研发领域, 被实验验证过的分子活性数据的数量往往非常有限。以ChEMBL为例, 多数靶点对应的化合物活性数据条目仅有数百至数千, 其规模低于一般训练深度学习模型所需的大数据的水平。除了数据数量外, 数据的多样性也是一个限制因素, 这导致AI模型可能在训练集和测试集上表现良好(因为训练集和测试集一般来自同一数据主体), 而应用于未知化合物库的虚拟筛选时, 其泛化能力不尽如人意, 出现较多的假阳性和假阴性结果。此外, 公共数据库中的实验数据通常在不同的方法和条件中进行测量^[126], 评价方法不统一, 减弱了数据的可靠性。为了应对这些挑战, 在数据供应方面, 还需要更多规范化的高通量筛选的数据。在数据使用方面, 则可以根据药学领域的相关理论, 与AI模型相结合, 或者对训练数据(尤其是非活性数据)进行合理扩充^[127-129]。此外, 联合其他的虚拟筛选方法, 比如分子对接、分子动力学和药效团模型等进行多角度评价, 也有望进一步减少AI模型的假阳性结果。

基于AI的药物筛选亟需合理的计算模型。AI的深度学习模型往往缺乏可解释性, 而人类在医学和生物学领域的很多领域认识并不透彻, 使得深度学习模型又常常被认为是“黑匣子”^[130], 它在实现预测方面缺乏明确性和透明度, 输入数据往往以一种无法解释的方式在多个隐藏层中进行复杂的转换^[131]。对于药物开发来说, 药物的作用机制几乎与药物效果同等重要, 这就对AI模型的可解释性提出了更高的要求。因此, AI模型的结果一般只能作为研究的起点, 还需要由研究人员进一步验证并展开研究。要提高模型的可解释性, 一方面要避免模型依赖于不相关变量, 另一方面还应当能够揭示潜在的生物学机制。目前, 一些研究正在努力打开人工神经网络模型的“黑匣子”^[132-136], 这方

面的研究有望进一步增加AI的可解释性。除可解释性外, AI模型的另一大隐患来自深度学习的“盲点”, 如通过对数据添加细微的噪声, 即可让神经网络的出错率大幅提升^[137], 这表明现有的AI技术还存在一些内生的缺陷, 使用AI去探索新的分子结构和未知的化学空间时, 也有可能陷入到这些盲点当中, 带来隐患。因此, 如何开发或选择合理的模型, 也是药物筛选中应当关注的问题。

5 展望

大数据将不断推进AI技术的应用。药物筛选需要高质量的生物活性数据, 随着分子生物学的发展, 必将有更多优质的数据被创造出来, 提升AI预测准确率。此外, 合理地从现有信息中挖掘数据可以创造AI应用的机会。Gkotsis等^[138]通过分析人们在网络平台上的发言, 利用深度学习自然语言处理技术, 预测其患有精神疾病的倾向。类似地, 在药学领域中, 每天都有大量的文献发表, 如果能够利用自然语言处理等技术从中合理地提取大数据, 必将能够创造出新的知识。

AI技术需要与现有的知识更紧密地结合。目前, 尽管AI技术不断取得突破, 但其定位仍然是数据处理的工具, 与真正的智能相去甚远。AI技术需要结合更广泛的知识来取得突破。如在药效预测方面, 目前的研究多局限于单成分单靶点的分析, 而从网络调控的角度入手的研究数量相对较少^[139,140]。再如, AI技术不仅可以从疾病寻找药物, 也可以从药物定位疾病。利用AI技术打破常规, 转换思维, 将会带来药物发现领域的突破。

随着AI技术的发展, 相信准确率更高、可解释性更强的AI方法将不断被开发出来。在AI的帮助下, 一定会有更多的新药能够低成本、高效率地被发明出来, 为人类的健康保驾护航。

作者贡献: 刘润哲撰写论文初稿; 宋俊科、刘艾林、杜冠华修改论文并定稿。

利益冲突: 所有作者均声明不存在利益冲突。

References

- [1] Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development [J]. *Nat Rev Drug Discov*, 2019, 18: 463-477.
- [2] Sirota M, Dudley JT, Kim J, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data [J]. *Sci Transl Med*, 2011, 3: 96ra77.
- [3] Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects [J]. *Drug Discov Today*, 2019, 24: 773-780.

- [4] Liu AL, Du GH. Research progress of virtual screening aided drug discovery [J]. *Acta Pharm Sin (药学报)*, 2009, 44: 566-570.
- [5] Russell SJ. *Artificial Intelligence: A Modern Approach* [M]. 3rd ed. Harlow: Pearson Education Limited Press, 2016: 1-2.
- [6] Samuel AL. Some studies in machine learning using the game of checkers [J]. *IBM J Res Dev*, 1959, 3: 210-229.
- [7] Rebalá G, Ravi A, Churiwala S. *An Introduction to Machine Learning* [M]. Cham: Springer Nature Switzerland AG, 2019: 1-3.
- [8] Mitchell TM. *Machine Learning* [M]. New York: McGraw-hill, 1997: 1-2.
- [9] Baum EB. On the capabilities of multilayer perceptrons [J]. *J Copmplexity*, 1988, 4: 193-215.
- [10] Pearl J. Bayesian networks: a model of self-activated memory for evidential reasoning [C]// *The Proceedings of the 7th Conference of the Cognitive Science Society*. Irvine: University of California, 1985: 15-17.
- [11] Cortes C, Vapnik V. Support-vector networks [J]. *Mach Learn*, 1995, 20: 273-297.
- [12] Quinlan JR. Simplifying decision trees [J]. *Int J Man Mach Stud*, 1987, 27: 221-234.
- [13] Yang X, Wang Y, Byrne R, et al. Concepts of artificial intelligence for computer-assisted drug discovery [J]. *Chem Rev*, 2019, 119: 10520-10594.
- [14] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python [J]. *J Mach Learn Res*, 2011, 12: 2825-2830.
- [15] Abadi M, Barham P, Chen J, et al. Tensorflow: a system for large-scale machine learning [C]// *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)*. Savannah: The Advanced Computing Systems Association, 2016: 265-283.
- [16] Ramsundar B, Eastman P, Walters P, et al. *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More* [M]. Sebastopol: O'Reilly Media, 2019.
- [17] Minnich AJ, McLoughlin K, Tse M, et al. AMPL: a data-driven modeling pipeline for drug discovery [J]. *J Chem Inf Model*, 2020, 60: 1955-1968.
- [18] Mangione W, Falls Z, Chopra G, et al. cando.py: open source software for predictive bioanalytics of large scale drug-protein-disease data [J]. *J Chem Inf Model*, 2020, 60: 4131-4136.
- [19] Thafar M, Raies AB, Albaradei S, et al. Comparison study of computational prediction tools for drug-target binding affinities [J]. *Front Chem*, 2019, 7: 782.
- [20] Jabeen A, Ranganathan S. Applications of machine learning in GPCR bioactive ligand discovery [J]. *Curr Opin Struct Biol*, 2019, 55: 66-76.
- [21] Achary PGR. Applications of quantitative structure-activity relationships (QSAR) based virtual screening in drug design: a review [J]. *Mini Rev Med Chem*, 2020, 20: 1375-1388.
- [22] Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery [J]. *Nucleic Acids Res*, 2012, 40: D1100-D1107.
- [23] Gaulton A, Hersey A, Nowotka M, et al. The ChEMBL database in 2017 [J]. *Nucleic Acids Res*, 2017, 45: D945-D954.
- [24] Wang Y, Bryant SH, Cheng T, et al. PubChem BioAssay: 2017 update [J]. *Nucleic Acids Res*, 2017, 45: D955-D963.
- [25] Wang Y, Xiao J, Suzek TO, et al. PubChem's bioassay database [J]. *Nucleic Acids Res*, 2012, 40: D400-D412.
- [26] Liu T, Lin Y, Wen X, et al. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities [J]. *Nucleic Acids Res*, 2007, 35: D198-D201.
- [27] Kumar R, Chaudhary K, Gupta S, et al. CancerDR: cancer drug resistance database [J]. *Sci Rep*, 2013, 3: 1445.
- [28] Ahmed J, Meinel T, Dunkel M, et al. CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge [J]. *Nucleic Acids Res*, 2011, 39: D960-D967.
- [29] Shankavaram UT, Varma S, Kane D, et al. CellMiner: a relational database and query tool for the NCI-60 cancer cell lines [J]. *BMC Genomics*, 2009, 10: 277.
- [30] Seiler KP, George GA, Happ MP, et al. ChemBank: a small-molecule screening and cheminformatics resource database [J]. *Nucleic Acids Res*, 2008, 36: D351-D359.
- [31] Lamb J, Crawford ED, Peck D, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease [J]. *Science*, 2006, 313: 1929-1935.
- [32] Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018 [J]. *Nucleic Acids Res*, 2018, 46: D1074-D1082.
- [33] Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: new perspectives on genomes, pathways, diseases and drugs [J]. *Nucleic Acids Res*, 2017, 45: D353-D361.
- [34] Koscielny G, An P, Carvalho-Silva D, et al. Open Targets: a platform for therapeutic target identification and validation [J]. *Nucleic Acids Res*, 2017, 45: D985-D994.
- [35] Cerami EG, Gross BE, Demir E, et al. Pathway Commons, a web resource for biological pathway data [J]. *Nucleic Acids Res*, 2011, 39: D685-D690.
- [36] Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database [J]. *Nucleic Acids Res*, 2002, 30: 412-415.
- [37] Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules [J]. *J Chem Inf Comput Sci*, 1988, 28: 31-36.
- [38] Todeschini R, Consonni V. *Molecular Descriptors for Chemoinformatics* [M]. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, 2009: 1-3.
- [39] Rogers D, Hahn M. Extended-connectivity fingerprints [J]. *J*

- Chem Inf Model, 2010, 50: 742-754.
- [40] Li P, Li Y, Hsieh CY, et al. TrimNet: learning molecular representation from triplet messages for biomedicine [J]. Brief Bioinform, 2021, 22: bbaa266.
- [41] O'Boyle NM, Banck M, James CA, et al. Open Babel: an open chemical toolbox [J]. J Cheminform, 2011, 3: 33.
- [42] Pahikkala T, Airola A, Pietila S, et al. Toward more realistic drug-target interaction predictions [J]. Brief Bioinform, 2015, 16: 325-337.
- [43] He T, Heidemeyer M, Ban F, et al. SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines [J]. J Cheminform, 2017, 9: 24.
- [44] Ozturk H, Ozgur A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction [J]. Bioinformatics, 2018, 34: i821-i829.
- [45] Rifaioğlu A, Sinoplu E, Atalay V, et al. DEEPScreen: high performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations [J]. Chem Sci, 2020, 11: 2531-2557.
- [46] Karimi M, Wu D, Wang Z, et al. DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks [J]. Bioinformatics, 2019, 35: 3329-3338.
- [47] Majumdar S, Nandi SK, Ghosal S, et al. Deep learning-based potential ligand prediction framework for COVID-19 with drug-target interaction model [J]. Cognit Comput, 2021. DOI: 10.1007/s12559-021-09840-x.
- [48] Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles [J]. Cell, 2017, 171: 1437-1452.e17.
- [49] Brum AM, van de Peppel J, van der Leije CS, et al. Connectivity map-based discovery of parabendazole reveals targetable human osteogenic pathway [J]. Proc Natl Acad Sci U S A, 2015, 112: 12711-12716.
- [50] van Noort V, Scholch S, Iskar M, et al. Novel drug candidates for the treatment of metastatic colorectal cancer through global inverse gene-expression profiling [J]. Cancer Res, 2014, 74: 5690-5699.
- [51] Dudley JT, Sirota M, Shenoy M, et al. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease [J]. Sci Transl Med, 2011, 3: 96ra76.
- [52] Chiu YC, Chen HH, Zhang T, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks [J]. BMC Med Genomics, 2019, 12: 18.
- [53] Aliper A, Plis S, Artemov A, et al. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data [J]. Mol Pharm, 2016, 13: 2524-2530.
- [54] Jeon M, Park D, Lee J, et al. ReSimNet: drug response similarity prediction using Siamese neural networks [J]. Bioinformatics, 2019, 35: 5249-5256.
- [55] Kim SK, Goughnour PC, Lee EJ, et al. Identification of drug combinations on the basis of machine learning to maximize anti-aging effects [J]. PLoS One, 2021, 16: e0246106.
- [56] Carrella D, Napolitano F, Rispoli R, et al. Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis [J]. Bioinformatics, 2014, 30: 1787-1788.
- [57] Xu H, Aldrich MC, Chen Q, et al. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality [J]. J Am Med Inform Assoc, 2015, 22: 179-191.
- [58] Kanjilal S, Oberst M, Boominathan S, et al. A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection [J]. Sci Transl Med, 2020, 12: eaay5067.
- [59] Scheeder C, Heigwer F, Boutros M. Machine learning and image-based profiling in drug discovery [J]. Curr Opin Syst Biol, 2018, 10: 43-52.
- [60] Onakpoya IJ, Heneghan CJ, Aronson JK. Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis [J]. Crit Rev Toxicol, 2016, 46: 477-489.
- [61] Segall MD, Barber C. Addressing toxicity risk when designing and selecting compounds in early drug discovery [J]. Drug Discov Today, 2014, 19: 688-693.
- [62] Ferreira LLG, Andricopulo AD. ADMET modeling approaches in drug discovery [J]. Drug Discov Today, 2019, 24: 1157-1165.
- [63] Ng HW, Doughty SW, Luo H, et al. Development and validation of decision forest model for estrogen receptor binding prediction of chemicals using large data sets [J]. Chem Res Toxicol, 2015, 28: 2343-2351.
- [64] Algamal Z, Qasim M, Ali H. A QSAR classification model for neuraminidase inhibitors of influenza A viruses (H1N1) based on weighted penalized support vector machine [J]. SAR QSAR Environ Res, 2017, 28: 415-426.
- [65] Wang NN, Dong J, Deng YH, et al. ADME properties evaluation in drug discovery: prediction of Caco-2 cell permeability using a combination of NSGA-II and boosting [J]. J Chem Inf Model, 2016, 56: 763-773.
- [66] Grenet I, Merlo K, Comet JP, et al. Stacked generalization with applicability domain outperforms simple QSAR on *in vitro* toxicological data [J]. J Chem Inf Model, 2019, 59: 1486-1496.
- [67] Ma J, Sheridan RP, Liaw A, et al. Deep neural nets as a method for quantitative structure-activity relationships [J]. J Chem Inf Model, 2015, 55: 263-274.
- [68] Basile AO, Yahi A, Tatonetti NP. Artificial intelligence for drug toxicity and safety [J]. Trends Pharmacol Sci, 2019, 40: 624-635.
- [69] Hu Q, Feng M, Lai L, et al. Prediction of drug-likeness using deep autoencoder neural networks [J]. Front Genet, 2018, 9: 585.
- [70] Altae-Tran H, Ramsundar B, Pappu AS, et al. Low data drug discovery with one-shot learning [J]. ACS Cent Sci, 2017, 3:

- 283-293.
- [71] Tian S, Djoumbou-Feunang Y, Greiner R, et al. CypReact: a software tool for *in silico* reactant prediction for human cytochrome P450 enzymes [J]. *J Chem Inf Model*, 2018, 58: 1282-1291.
- [72] Lagorce D, Bouslama L, Becot J, et al. FAF-Drugs4: free ADME-tox filtering computations for chemical biology and early stages drug discovery [J]. *Bioinformatics*, 2017, 33: 3658-3660.
- [73] Podlewska S, Kafel R. MetStabOn-online platform for metabolic stability predictions [J]. *Int J Mol Sci*, 2018, 19: 1040.
- [74] Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules [J]. *Sci Rep*, 2017, 7: 42717.
- [75] Stork C, Chen Y, Šicho M, et al. Hit Dexter 2.0: machine-learning models for the prediction of frequent hitters [J]. *J Chem Inf Model*, 2019, 59: 1030-1043.
- [76] Schyman P, Liu R, Desai V, et al. vNN Web server for ADMET predictions [J]. *Front Pharmacol*, 2017, 8: 889.
- [77] Dong J, Wang NN, Yao ZJ, et al. ADMETlab: a platform for systematic ADMET evaluation based on a comprehensively collected ADMET database [J]. *J Cheminform*, 2018, 10: 29.
- [78] Klambauer G, Unterthiner T, Mayr A, et al. DeepTox: toxicity prediction using deep learning [J]. *Toxicol Lett*, 2017, 280: S69.
- [79] Peng Y, Zhang Z, Jiang Q, et al. TOP: a deep mixture representation learning method for boosting molecular toxicity prediction [J]. *Methods*, 2020, 179: 55-64.
- [80] Hyunho K, Hojung N. hERG-Att: self-attention-based deep neural network for predicting hERG blockers [J]. *Comput Biol Chem*, 2020, 87: 107286.
- [81] Paranjpe MD, Taubes A, Sirota M. Insights into computational drug repurposing for neurodegenerative disease [J]. *Trends Pharmacol Sci*, 2019, 40: 565-576.
- [82] Stumpfe D, Bajorath J. Exploring activity cliffs in medicinal chemistry [J]. *J Med Chem*, 2012, 55: 2932-2942.
- [83] Sledz P, Caflisch A. Protein structure-based drug design: from docking to molecular dynamics [J]. *Curr Opin Struct Biol*, 2018, 48: 93-102.
- [84] Chen YC. Beware of docking! [J]. *Trends Pharmacol Sci*, 2015, 36: 78-95.
- [85] Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions [J]. *Phys Chem Chem Phys*, 2010, 12: 12899-12908.
- [86] Copeland RA. The dynamics of drug-target interactions: drug-target residence time and its impact on efficacy and safety [J]. *Expert Opin Drug Discov*, 2010, 5: 305-310.
- [87] Li H, Peng J, Sidorov P, et al. Classical scoring functions for docking are unable to exploit large volumes of structural and interaction data [J]. *Bioinformatics*, 2019, 35: 3989-3995.
- [88] Fresnais L, Ballester PJ. The impact of compound library size on the performance of scoring functions for structure-based virtual screening [J]. *Brief Bioinform*, 2020. DOI: 10.1093/bib/bbaa095.
- [89] Li H, Leung KS, Wong MH, et al. Improving AutoDock Vina using random forest: the growing accuracy of binding affinity prediction by the effective exploitation of larger data sets [J]. *Mol Inform*, 2015, 34: 115-126.
- [90] Zilian D, Sotriffer CA. SFCscore RF: a random forest-based scoring function for improved affinity prediction of protein-ligand complexes [J]. *J Chem Inf Model*, 2013, 53: 1923-1933.
- [91] Li H, Leung KS, Ballester PJ, et al. istar: a web platform for large-scale protein-ligand docking [J]. *PLoS One*, 2014, 9: e85678.
- [92] Torng W, Altman RB. Graph convolutional neural networks for predicting drug-target interactions [J]. *J Chem Inf Model*, 2019, 59: 4131-4149.
- [93] Skalic M, Martinez-Rosell G, Jimenez J, et al. PlayMolecule BindScope: large scale CNN-based virtual screening on the web [J]. *Bioinformatics*, 2019, 35: 1237-1238.
- [94] Jiménez J, Škalič M, Martínez-Rosell G, et al. K(DEEP): protein-ligand absolute binding affinity prediction *via* 3D-convolutional neural networks [J]. *J Chem Inf Model*, 2018, 58: 287-296.
- [95] Jiménez-Luna J, Pérez-Benito L, Martínez-Rosell G, et al. DeltaDelta neural networks for lead optimization of small molecule potency [J]. *Chem Sci*, 2019, 10: 10911-10918.
- [96] Spitzer R, Jain AN. Surfex-Dock: docking benchmarks and real-world application [J]. *J Comput Aided Mol Des*, 2012, 26: 687-699.
- [97] Allen WJ, Balias TE, Mukherjee S, et al. DOCK 6: impact of new features and current docking performance [J]. *J Comput Chem*, 2015, 36: 1132-1156.
- [98] Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning [J]. *Nature*, 2020, 577: 706-710.
- [99] AlQuraishi M. End-to-end differentiable learning of protein structure [J]. *Cell Syst*, 2019, 8: 292-301.e3.
- [100] Kana O, Brylinski M. Elucidating the druggability of the human proteome with eFindSite [J]. *J Comput Aided Mol Des*, 2019, 33: 509-519.
- [101] Yuan JH, Han SB, Richter S, et al. Druggability assessment in TRAPP using machine learning approaches [J]. *J Chem Inf Model*, 2020, 60: 1685-1699.
- [102] Bakheet TM, Doig AJ. Properties and identification of human protein drug targets [J]. *Bioinformatics*, 2009, 25: 451-457.
- [103] Kim B, Jo J, Han J, et al. *In silico* re-identification of properties of drug target proteins [J]. *BMC Bioinformatics*, 2017, 18: 248.
- [104] Wang Q, Feng Y, Huang J, et al. A novel framework for the identification of drug target proteins: combining stacked auto-encoders with a biased support vector machine [J]. *PLoS One*, 2017, 12: e0176486.
- [105] Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Improving detection of protein-ligand binding sites with 3D segmentation [J]. *Sci Rep*, 2020, 10: 5035.

- [106] Nayal M, Honig B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites [J]. *Proteins*, 2006, 63: 892-906.
- [107] Chan HS, Li Y, Dahoun T, et al. New binding sites, new opportunities for GPCR drug discovery [J]. *Trends Biochem Sci*, 2019, 44: 312-330.
- [108] Wu Q, Peng Z, Zhang Y, et al. COACH-D: improved protein-ligand binding sites prediction with refined ligand-binding poses through molecular docking [J]. *Nucleic Acids Res*, 2018, 46: W438-W442.
- [109] Olivecrona M, Blaschke T, Engkvist O, et al. Molecular *de-novo* design through deep reinforcement learning [J]. *J Cheminform*, 2017, 9: 48.
- [110] Liu X, Ye K, van Vlijmen HWT, et al. An exploration strategy improves the diversity of *de novo* ligands using deep reinforcement learning: a case for the adenosine A(2A) receptor [J]. *J Cheminform*, 2019, 11: 35.
- [111] Zhou Z, Kearnes S, Li L, et al. Optimization of molecules *via* deep reinforcement learning [J]. *Sci Rep*, 2019, 9: 10752.
- [112] Segler MH, Kogej T, Tyrchan C, et al. Generating focused molecule libraries for drug discovery with recurrent neural networks [J]. *ACS Cent Sci*, 2018, 4: 120-131.
- [113] Gupta A, Müller AT, Huisman BJ, et al. Generative recurrent networks for *de novo* drug design [J]. *Mol Inform*, 2018, 37: 1700111.
- [114] Putin E, Asadulaev A, Ivanenkov Y, et al. Reinforced adversarial neural computer for *de novo* molecular design [J]. *J Chem Inf Model*, 2018, 58: 1194-1204.
- [115] Bian Y, Wang J, Jun JJ, et al. Deep convolutional generative adversarial network (dcGAN) models for screening and design of small molecules targeting cannabinoid receptors [J]. *Mol Pharm*, 2019, 16: 4451-4460.
- [116] Kadurin A, Nikolenko S, Khrabrov K, et al. druGAN: an advanced generative adversarial autoencoder model for *de novo* generation of new molecules with desired molecular properties *in silico* [J]. *Mol Pharm*, 2017, 14: 3098-3104.
- [117] Méndez-Lucio O, Baillif B, Clevert DA, et al. *De novo* generation of hit-like molecules from gene expression signatures using artificial intelligence [J]. *Nat Commun*, 2020, 11: 10.
- [118] Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules [J]. *ACS Cent Sci*, 2018, 4: 268-276.
- [119] Kang S, Cho K. Conditional molecular design with deep generative models [J]. *J Chem Inf Model*, 2019, 59: 43-52.
- [120] Skalic M, Jiménez J, Sabbadin D, et al. Shape-based generative modeling for *de novo* drug design [J]. *J Chem Inf Model*, 2019, 59: 1205-1214.
- [121] Merk D, Friedrich L, Grisoni F, et al. *De novo* design of bioactive small molecules by artificial intelligence [J]. *Mol Inform*, 2018, 37: 1700153.
- [122] Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors [J]. *Nat Biotechnol*, 2019, 37: 1038-1040.
- [123] Yang Y, Zhang R, Li Z, et al. Discovery of highly potent, selective, and orally efficacious p300/CBP histone acetyltransferases inhibitors [J]. *J Med Chem*, 2020, 63: 1337-1360.
- [124] Blaschke T, Arús-Pous J, Chen H, et al. REINVENT 2.0: an AI tool for *de novo* drug design [J]. *J Chem Inf Model*, 2020, 60: 5918-5922.
- [125] Skalic M, Sabbadin D, Sattarov B, et al. From target to drug: generative modeling for the multimodal structure-based ligand design [J]. *Mol Pharm*, 2019, 16: 4282-4291.
- [126] Davies M, Nowotka M, Papadatos G, et al. ChEMBL web services: streamlining access to drug discovery data and utilities [J]. *Nucleic Acids Res*, 2015, 43: W612-W620.
- [127] Xia J, Jin H, Liu Z, et al. An unbiased method to build benchmarking sets for ligand-based virtual screening and its application to GPCRs [J]. *J Chem Inf Model*, 2014, 54: 1433-1450.
- [128] Mysinger MM, Carchia M, Irwin JJ, et al. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking [J]. *J Med Chem*, 2012, 55: 6582-6594.
- [129] Lee AA, Yang Q, Bassyouni A, et al. Ligand biological activity predicted by cleaning positive and negative chemical correlations [J]. *Proc Natl Acad Sci U S A*, 2019, 116: 3373-3378.
- [130] Gilvary C, Madhukar N, Elkhader J, et al. The missing pieces of artificial intelligence in medicine [J]. *Trends Pharmacol Sci*, 2019, 40: 555-564.
- [131] Lee CY, Chen YP. Machine learning on adverse drug reactions for pharmacovigilance [J]. *Drug Discov Today*, 2019, 24: 1332-1343.
- [132] Zhang Q, Nian Wu Y, Zhu SC. Interpretable convolutional neural networks [C] // *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2018: 8827-8836.
- [133] Sheridan RP. Interpretation of QSAR models by coloring atoms according to changes in predicted activity: how robust is it? [J]. *J Chem Inf Model*, 2019, 59: 1324-1337.
- [134] Weibel HE, Kimber TB, Radetzki S, et al. Revealing cytotoxic substructures in molecules using deep learning [J]. *J Comput Aided Mol Des*, 2020, 34: 731-746.
- [135] Samek W. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* [M]. Cham: Springer Nature Switzerland AG, 2019: 51-53.
- [136] Kuenzi BM, Park J, Fong SH, et al. Predicting drug response and synergy using a deep learning model of human cancer cells [J]. *Cancer Cell*, 2020, 38: 672-684.e6.
- [137] Yuan X, He P, Zhu Q, et al. Adversarial examples: attacks and defenses for deep learning [J]. *IEEE Trans Neural Netw Learn Syst*, 2017, 30: 2805-2824.
- [138] Gkotsis G, Oellrich A, Velupillai S, et al. Characterisation of

- mental health conditions in social media using informed deep learning [J]. *Sci Rep*, 2017, 7: 45141.
- [139] Hou Y, Nie Y, Cheng B, et al. Qingfei Xiaoyan Wan, a traditional Chinese medicine formula, ameliorates *Pseudomonas aeruginosa*-induced acute lung inflammation by regulation of PI3K/AKT and Ras/MAPK pathways [J]. *Acta Pharm Sin B*, 2016, 6: 212-221.
- [140] Liu AL, Du GH. Network pharmacology: new guidelines for drug discovery [J]. *Acta Pharm Sin (药学报)*, 2010, 45: 1472-1477.