

DOI:10.3969/j.issn.1005-202X.2024.11.017

医学人工智能

基于人工智能技术驱动的新药研发

岑菲菲¹, 陈良²

1. 乐山职业技术学院生物医药学院, 四川 乐山 614000; 2. 成都倍特药业股份有限公司, 四川 成都 610095

【摘要】通过深度学习、机器学习等技术,能够迅速地筛选潜在的药物候选分子,并预测其与生物靶标的相互作用。此外,人工智能在药物设计、合成路径规划、临床试验数据分析等方面展现出巨大潜力,有望缩短新药从实验室到市场的时间,降低研发成本。本文综述当前人工智能技术在新药研发中的应用进展,探讨其面临的挑战与未来发展方向。

【关键词】人工智能技术;深度学习;新药研发;综述

【中图分类号】R318;R913

【文献标志码】A

【文章编号】1005-202X(2024)11-1437-06

New drug research and development driven by artificial intelligence technology

CEN Feifei¹, CHEN Liang²

1. School of Biomedicine, Leshan Vocational and Technical College, Leshan 614000, China; 2. Chengdu Brilliant Pharmaceutical Co., Ltd., Chengdu 610095, China

Abstract: The technologies such as deep learning and machine learning enable the rapid screening for potential drug candidate molecules and the prediction of their interactions with biological targets. Additionally, artificial intelligence exhibits great potential in drug design, synthesis pathway planning and clinical trial data analysis, and is expected to shorten the time from laboratory to market for new drugs and reduce research and development costs. Herein the study reviews the application of artificial intelligence technology in new drug research and development, explores the challenges it faces, and discusses the future development directions.

Keywords: artificial intelligence technology; deep learning; new drug research and development; review

前言

在过去的数十年中,药物研发一直是一个耗时且成本高昂的过程,涉及复杂的生物化学实验、庞大的数据分析以及长期的临床试验。然而,随着人工智能(AI)的不断进步,这一领域正在经历一场深刻的变革。AI在新药研发中的应用体现在通过深度学习和机器学习算法,可以对大量的生物医学文献和实验数据进行快速分析,识别出与疾病相关的生物标志物和潜在的药物靶点^[1]。同时,AI能够模拟药物与靶点的相互作用,预测药物的活性和毒性,为药物设计提供理论依据。总之,AI为新药研发带来了前所未有的机遇,它不仅能够显著缩短药物研发周期,

还能够提高新药研发的成功率和质量^[2]。

1 新药研发的基本流程

随着现代生物医学理论的持续演变,现代生物技术的持续发展,药物的发现在技术上可以划分为:药物作用靶点的选择和测定,生物标志物、先导化合物的测定和候选药物的测定。临床前开发阶段以筛选活性化合物和研究构效关系为主,包括药物疗效分析、安全性评估和药物动力学参数评估^[3]。在临床阶段,主要包括药物重定向、患者招募和临床试验,包括药物筛选、药物有效性优化等。审批和上市由医药公司的研发部门配合国家药品监管部门完成。一家医药公司通常需投资数千万至数亿美元研发新药,整个过程耗时超过10年^[4]。但是,在漫长的时间里,经历了无数的考验,最终成功上市的药物却少之又少。经过调查发现,在2011~2020年这段时间里,药物从1期临床到FDA批准上市,成功率平均为7.9%,所需时间平均10.5年。尽管各个医药公司的

【收稿日期】2024-04-14

【基金项目】四川省教育厅教育科研课题(SCJG21A257);乐山市科技局重点研究项目(19SZD114)

【作者简介】岑菲菲,硕士,讲师,主管药师,研究方向:人工智能在药理学、临床药学、药物分析的应用, E-mail: feifei-20150806@163.com

营业收入都有不同程度的提高,但是却无法跟上研发费用的增加。随着投资回报率的降低,新药研发的成功率也在不断降低。但随着新药智能研发的开始,这个现象可能会有所改变^[5]。

2 AI+新药研发的应用模式

2.1 AI+新药研发应用算法

AI常见算法见表1,包括机器学习(ML)和深度学习(DL)等。其中,ML算法被广泛地应用于药物研究与开发中,如决策树、随机森林、支持向量机、k-最近邻算法、朴素贝叶斯分类器等;DL算法主要包含深层神经网络(DNN)、卷积神经网络(CNN)、循环神经网络(RNN)、自编码器等,适用于大规模的数据集,具有较高的计算复杂度。随着计算机处理能力的不断

提升,以及海量数据的不断积累,DL算法已被广泛地应用于新药开发领域。DNN是最早用于新药研发的一种深度学习算法,其起源于1943年McCulloch等^[6]提出的一种计算模型。CNN作为一种前向神经网络,在图像识别方面表现突出。RNN是一种处理序列的神经网络,可以对序列数据进行处理,并具有一定的记忆功能。自编码器主要用于对输入数据进行重建,并具有一定的学习能力,因此在药物合成领域具有广泛的应用前景。DNN、CNN、RNN等DL算法模型通过定量结构性质关系(QSPR)或定量结构活性关系(QSAR)等预测药物分子的物理化学性质以及药物在体内的吸收、分布、代谢、排泄以及毒性等^[7]。本文具体综述DL算法在新药研发中的应用。

表1 AI常见应用算法

Table 1 Common application algorithms of artificial intelligence

AI技术	算法类型	算法特点
机器学习(ML)	决策树	可以自动选择最佳特征进行划分,并生成可视化的决策树模型,使决策过程更加透明和可解释,适用于处理大规模数据集
	随机森林	由多个决策树组成,能够处理大量的特征和样本,具有很好的鲁棒性和泛化能力,在解决分类和回归问题时表现出色
	支持向量机	通过将数据映射到高维空间,可以找到一个最优超平面,将不同类别的数据点分开。同时,支持向量机还能够通过核函数的使用,处理非线性的分类问题
	k-最近邻算法	通过计算待分类样本与训练集中最近的k个样本的距离来确定其类别,这个算法不需要任何训练过程,只需要存储训练样本
	朴素贝叶斯分类器	朴素贝叶斯分类器算法计算简单、效率高,尤其适用于处理高维度数据。通过计算每个类别的概率,并选择具有最高概率的类别作为预测结果。朴素贝叶斯分类器在文本分类、垃圾邮件过滤等领域被广泛应用
深度学习(DL)	深度神经网络(DNN)	DNN通过学习输入数据的特征和模式,能够自动提取高级抽象表示,在图像识别、语音识别和自然语言处理等领域取得了显著成果
	循环神经网络(RNN)	RNN的特点在于它的隐藏层是循环连接的,循环结构使得RNN能够捕捉到序列中的上下文信息,并在处理长序列时具有优势。此外,RNN还可以通过反向传播算法进行训练,使得它能够适应各种任务,如语言模型、机器翻译和情感分析等
	卷积神经网络(CNN)	CNN具有局部感知性,可以捕捉到图像中的局部模式,同时具有参数共享和空间不变性的特点,使其在图像识别和计算机视觉任务中表现出色
	自编码器算法(AE)	AE是在无监督学习中使用了编码器和解码器,可以通过将输入数据压缩为低维编码,然后再将其解码回原始形式来重建数据

2.2 DL

以靶点为基础的药物研发,往往要借助打分函数对小分子与蛋白质的结合强度进行预测^[8]。传统打分函数包括基于力场、经验和知识的3种类型。DL通过提取三维空间结构信息预测活性。目前,模型间存在显著差异,主要由于训练集和表征方式的不同。训练集分为实验获取和分子对接预测的蛋白-配体复合体结构两类。前者以PDB Bind数据库为

代表,数据更可靠,但是数据量小;后者可以产生大量结构数据用于训练,但依赖于分子对接的准确度,训练集中混入错误的配体结合构象会影响DL的表现。表征方式主要有两种:一种是基于3D格点技术,将蛋白质-配体复合物视为3D图像,利用3D卷积模型进行蛋白质-配体相互作用模式的鉴定;二是将相邻的原子环境编码成一个原子环境矢量,用来描述蛋白质与配基的相互作用。前者有计算机视觉方面

的现成软件,可以用来进行前期的研究,而后者则是建立在原子产生特性的基础上,可以更好地刻画原子间的相互作用。对于三维格点的特性,人们已经有了比较多的认识,而对于原子尺度上的卷积,人们也在不断探索中^[9]。2007年,Behler等^[10]开创性利用高斯径向函数和高斯角函数作为基组编码原子位置信息,引入神经网络表征量化计算密度泛函理论,从而实现了与DFT相比的快速运算。2017年,ANI-1又采用了修正后的2007年Behler和Parrinello的对称函数,构建了一个单原子环境向量(AEV)来表征分子,因此,DL可以在DFT量化计算的结果基础上进行训练,学习生成的ANI-1能够计算出一个比训练集所包含的系统更大的系统,并且与DFT的定量计算精度相当,但是速度更快,这对于快速评估有机小分子的能量有很大帮助^[11]。就像DUD是评价分子对接算法的标准集一样,MoleculeNet将现有的数据集、表征方法和网络模型放入Github DeepChem中,并将其作为标准(Benchmark)测试^[12]。

3 基于DL的新药研发

3.1 新药研发算法基本框架

该方法可划分为4大类,包括编码器-解码器(Enc-Dec)、RNN、产生对抗网络(GAN)和强化学习(RL)。在过去的5年中,已开发出近100个基于DL的新药开发模型。这些模型的改进主要体现在两个方面:分子表征和多目标优化策略^[13]。基于DL框架的新药开发模型能够采用多种分子表示形式作为输入,包括简化分子输入线测定(SMILES)、分子指纹(MFP)、二维(2D)或三维(3D)分子图像等。多目标优化是同时使用不同的策略优化生成的几个分子的属性,如使用不同的DL框架限制和控制生产阶段允许分子在高度可堆肥的同时保持每个特定目标。目前,研究人员已经成功地使用DL框架设计和合成了活性苗头化合物,这比传统的新药设计方法花费的时间要短得多,并且具有很大的潜力^[14]。

3.2 新药研发模型中使用的DL框架

3.2.1 Enc-Dec 框架 Enc-Dec是一个无监督的DL框架,它包含编码和解码两个部分,编码部分将被编码的分子特征映射为一个带有隐性的图层,而解码部分将隐性图层的矢量映射为原始图层^[15]。变分自编码器(VAE)是产生式模式中应用最为广泛的一种Enc-Dec架构。VAE模型的训练主要是为了减少所产生的分子特征和原来的分子特征之间的偏差,如式(1)所示:

$$\max_{\theta} \int q(z) \log p(x|z) dz - \text{KL}(q(z|x) || p(z)) \quad (1)$$

训练过程主要是通过最大下限(LB)与编码器 $q(z|x)$ 与译码器 $p(x|z)$ (KL为KL散度,此处 $p(z)$ 为高斯分布)的关系式来进行。Chem VAE是国际上最早提出的一种新型药物研发方法,该方法通过对多层次感受器的预测,使隐藏层Z内的分子按照不同的性质(类药性、合成性等)呈现出不同程度的梯度分布,从而实现对样本数据的有效调控。

随后,许多研究人员开发了GVAE、JTVAE、SDVAE和CVAE等VAE模型,这些模型通常使用不同的分子结构,并使用CFG作为分子结构添加了SDVAE、GVAE等实用功能,可以使隐藏层中相邻分子的结构相似性更紧密,其中随机采样产生更灵敏的分子。在相同的数据训练中,SDVAE、GVAE和JTVAE产生的化合物在结构上明显比Chem VAE生产的化合物更敏感和新颖。CGVAE、Shape VAE、JTVAE和Graph VAE表征MG形式的分子。与SMILES相比,这些表征方法的概率分布更适合VAE模型的连续延迟层,并且与CFG一样,隐藏层中的分子结构分布更平滑,提高了采样的合理性和新颖性。CVAE和SSVAE在药物设计任务中为VAE框架添加了实用功能。用户可以使用预定义的骨架或属性值自定义生成的分子。尽管这些模型在一定程度上提高了分子生产的效率和新颖性,但它们的基本框架和原理没有发生显著变化。

除了VAE之外,在药物研究开发过程中,还引入了一种基于数据转化器的新型药物研究方法。这种方法通过DL模型,对大量药物分子数据进行学习和转化,从而发现新的药物候选分子。研究人员可以利用这种方法,对药物分子的结构、性质以及作用机制进行深入分析,从而推动药物研发的进程^[16]。该方法与传统的基于CNN和RNN的Enc-Dec产生方法有很大区别:首先,该方法采用了注意力机制来嵌入分子表征的词汇量。其次,在训练结束后,数据转化器(Transformer)将按照顺序和自回归的方法输出SMILES特征,并将其作为下一个SMILES特征的依据,使用Masked多头自动注意层,避免序列中未译码区域的词汇量信息泄漏。Transformer的核心是多头自注意层,这个模型包含了多个可扩展的点积注意力函数,使得模型能够依序获得关键信息,注意机制的计算步骤如式(2):

$$\text{attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V} \quad (2)$$

其中, \mathbf{Q} 、 \mathbf{K} 和 \mathbf{V} 分别表示查询矩阵、键矩阵和值矩阵, d_k 表示 \mathbf{K} 的维数。

3.2.2 RNN 框架 RNN 模型采用与文本相似的方法产生分子,以分子中的原子和片段为基础,按时间顺序产生分子。具有相同 DL 框架的模型总体性能相似,使用不同框架的模型性能指标差异较大。例如基于 RNN 模型的一些指标通常是优秀的,而基于 GAN 模型则相对较差^[17]。RNN 模型产生下一个词汇的概率取决于前面几层的结果,该训练过程如式(3)所示:

$$\max p(x) = p(s_1) \prod_{T=0}^T P(s_T | s_{T-1}, \dots, s_1) P \quad (3)$$

由此可见,RNN 分子生成模型试图最大化分子字符串 (y_1, y_2, \dots, y_i) 的有效组合的概率,从而以最大的概率生成下一个词汇。为了解决梯度衰减问题,提出使用具有长短期记忆的人工神经网络(LSTM)或其简化形式的门控神经网络(GRU)。迁移学习(TL)被广泛应用于 RNN 中,它的含义是将已有的研究成果转化为新的研究内容。目前已有大量的研究利用 TL 来精细调控 RNN 模型,产生抑制小分子。与其它 DL 方法相比,RNN 方法的优点是生成分子的长度没有任何限制,并且模型的建立与训练更加容易。

3.2.3 RL 框架 RL 是用于解决药物开发过程中的动态决策问题,它采用一种奖励函数(通常是基于机器学习的属性预测器),通过对生成算法中某一分子的动作策略 S_T 进行打分,并将打分结果反馈给生成算法,用于指导药物开发过程中的决策。式(4)给出 RL 的训练过程:

$$\max R = \sum_{s_T \in S} P(s_T) r(s_T) \quad (4)$$

该算法的训练方法主要是采用预测器对各个状态 S_T 进行奖赏,使得终端状态 S 的总分 R 最大。REINFORCE 和蒙特卡洛树搜索(MCTS)都是 RL 模型常用的策略函数。比如,在 REINVENT 产生模型中,采用修正的 REINFORCE 策略函数,在保证产生的分子正确性的同时,兼顾产生的分子物理化学特性与训练数据的一致性。在此基础上,利用 REINVENT 对药物进行结构修饰,并对其进行构效关系研究。接着,设计出一种新的 D2 受体(DRD2)拮抗剂,并证实这种模式可以设计出针对特定目标的药物。Deep FMPO 还采用 REINFORCE 策略功能,将 DRD2 和 DRD4 的活性分子数据转化为片段库,从而根据片段而非原子生成分子^[18]。最后制造出一组分子,这些分子的物理化学特性类似于 DRD2 和 DRD4。

MCTS 策略功能能够对当前部分生成的片段以及后续全部生成的片段进行评估。Chem TS 是一种结合了 RNN 与 MCTS 策略函数的生成模型。在此基

础上,将其与 RNN、Chem VAE、Grammar VAE 进行对比,结果表明该方法能够以更小的计算量得到更好的物理化学性质。在此基础上,Drug EX 将 MCTS 与双向 RNN 相结合,在同一发生器中融合了两种不同的 RNN,以确保产生的分子具有多样性和新颖性^[19]。

3.2.4 GAN 框架 GAN 模型由生成器 G 和鉴别器 D (也可以是属性预测器)组成,如式(5)所示:

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (5)$$

原理是生成器和鉴别器之间的零和博弈。鉴别器试图将发生器产生的分子识别为假 ($G(D(x)) \rightarrow 0, D(G(z)) \rightarrow 0$),而发生器试图从噪声中产生分子来“背叛”鉴别器 $D(G(z)) \rightarrow 1$ 。在新药开发模型中,GAN 通常与其他 DL 框架结合使用。例如,训练过程中产生的分子通过结合 RL、ORGAN 和 ORGAN 分离器和 RL 奖励系统进行评分,还可以与 VAE 相结合,对抗自动编码器(AAE)融合算法。AAE 通过引入影响 VAE 潜在层的 GAN 分离器,能够生产用于特定目标取样的化合物。与 VAE 相比,AAE 更适合表征分子 MFP 作为框架的输入。

4 新药研发模型预测

4.1 新分子的生成及活性预测

RNN 以序列信息为输入特征,曾被广泛地应用于自然语言处理,目前已被应用于新化合物结构的构建。Segler 等^[20]使用常用的 SMILES 字符串格式来表征化合物,RNN 首先通过类似学习语言的方式,学习了大量的 SMILES 文本是如何表征分子的,从而建立出的模型能够生成全新的 SMILES 字符串,也就是没有偏向性的全新分子,可以应用于虚拟筛选等多种用途。其次,通过迁移学习,可以在不需要海量数据的情况下,利用特定目标化合物的生物信息学数据,对已有的生物信息学数据进行重新构建。在此基础上,通过优化模型,实现以两种致病菌为靶标的新型药物开发。Reker 等^[21]利用前期研发的 SPiDER 技术,以 RXR 和 PPARs 为蛋白靶标,利用 DL 技术,成功合成了 5 个结构新颖全新分子,并发现 4 个具有从纳摩尔到低微摩尔范围内活性的化合物。Olivecrona 等^[22]进一步引入强化学习调整上述的 RNN 模型,从而保持总体一致性,避免生成的分子逐渐偏离方向和重复生成相同分子。配体分子对(MMP)分析是新药研发的重要手段,不同结构单元之间的不同位置往往会导致活性、溶解性、生物利用度、毒性等性能的改变。Turk 等^[23]将 MMP 与 DL 结合起来,使用逆合成分析的规则判断和提取 EMBL

数据库中的匹配分子对作为数据集,使用DNN训练匹配分子对活性发生变化的模式,然后根据得到的模型生成新的分子,并对其活性变化进行预测。

4.2 代谢及毒理预测

定量结构性质长期以来一直被用于预测成分的毒性,经验规则也被广泛使用。例如ToxAlerts数据库,该数据库收集了大量据报道会引起副作用的分子或分子结构,表明类似的分子可以引起相同或相似的副作用。DL也被研究并应用于预测化合物的代谢和毒理学特性。Fraser等^[24]基于DL开发了一种用于预测化合物不良反应的方法。Tox21数据预测不良反应的准确性与当前其他机器学习方法相当^[25]。由于药物毒性通常来自体内药物的活性代谢产物,如药物诱导的肝毒性。Hughes等^[26]开发了第一个DL模型预测化合物是否通过形成单级或两级醌和潜在毒性,并预测化合物是否形成环氧化物并可能导致毒性,可用于快速筛选化合物的潜在毒性。同时,DL方法也被用于预测化合物的其他性质。例如,Lusci等^[27]开发了AquaSol,通过改进化合物的表征并将其与DL-RNN相结合预测化合物的溶解度,其精度与当前的计算方法相当。

4.3 化合物合成路线预测

当前,化学合成路径的设计主要依赖于化学工作者自身的经验,而应用最为广泛的方法就是由Corey^[28]提出的逆向合成。IBM公司利用大量的专利和文献,对化学反应进行了研究,并对其进行了预测,这将化学反应的预测变成了语言翻译的问题^[29]。Segler等^[30]将DL与蒙特卡罗算法相结合,学习了大量经过多次验证的化学反应,然后使用反向合成策略设计合成路线并选择合适的前体。这种方法比常用的计算机辅助合成路线设计方法更有效,并在专业的双盲测试中得到了认可。

5 小结

通过结合AI中的DL算法,新药研发迎来了机遇。DL在药物筛选、设计和优化等方面展现了巨大的潜力,为研发人员提供了强大的工具和方法,其高效性和精确性有望大幅降低研发时间和成本,并提高药物研发的成功率。

【参考文献】

[1] 张婷,陈娟,欧阳昭连. 基于专利分析的人工智能辅助药物研发领域技术开发态势[J]. 中国新药杂志, 2022, 31(4): 305-311.
Zhang T, Chen J, Ouyang ZL. The technology development trends in the field of artificial intelligence-assisted drug discovery based on patent analysis [J]. Chinese Journal of New Drugs, 2022, 31(4): 305-311.

[2] 杨洪军. 改良型新药: 中药大品种培育新方向[J]. 中国药理学与毒

理学杂志, 2021, 35(10): 726.

Yang HJ. Improved new drugs: a new direction for the cultivation of large varieties of traditional Chinese medicine [J]. Chinese Journal of Pharmacology and Toxicology, 2021, 35(10): 726.

[3] 乔连生,李军,谢兰,等. 基于靶向转录组、专家经验和人工智能研发守正创新中药的路径探索-以治疗慢性心力衰竭的中药新药研发为例[J]. 中医杂志, 2023, 64(3): 217-224.
Qiao LS, Li J, Xie L, et al. The exploration of a new approach for developing innovative Chinese herbal formulae by combining targeted transcriptome, expert experience and artificial intelligence: taking the development of anti-chronic heart failure innovative formula as an example [J]. Journal of Traditional Chinese Medicine, 2023, 64(3): 217-224.

[4] 张志豪,周吟玦,姜慧君,等. 人工智能药物发现研究体系的构建及实践[J]. 中国新药杂志, 2022, 31(13): 1294-1303.
Zhang ZH, Zhou YJ, Jiang HJ, et al. Construction and practice of artificial intelligence in drug discovery system [J]. Chinese Journal of New Drugs, 2022, 31(13): 1294-1303.

[5] 赵锐,石秀园,钟雪然,等. 基于卫生技术评估构建医学人工智能技术临床应用评估路径的探讨[J]. 中国卫生政策研究, 2022, 15(2): 78-82.
Zhao R, Shi XY, Zhong XR, et al. Enoploring of construction of clinical application assessment path on medical artificial intelligence technology based on health technology assessment [J]. Chinese Journal of Health Policy, 2022, 15(2): 78-82.

[6] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity [J]. Bull Math Biol, 1943, 5(4): 115-133.

[7] 刘昌孝,张铁军. 基于“物质-药代-功效”关联的中药创新研发思路[J]. 中草药, 2022, 53(1): 1-7.
Liu CX, Zhang TJ. Innovative research and development ideas of traditional Chinese medicine based on correlation of substance-pharmacokinetics-effects [J]. Chinese Traditional and Herbal Drugs, 2022, 53(1): 1-7.

[8] 谷书凯,王哲,侯廷军,等. 分子动力学模拟和机器学习结合在药物设计领域的应用[J]. 中国现代应用药学, 2022, 39(21): 2804-2808.
Gu SK, Wang Z, Hou TJ, et al. Application of molecular dynamics simulations meet machine learning in structure-based drug design [J]. Chinese Journal of Modern Applied Pharmacy, 2022, 39(21): 2804-2808.

[9] 茅鸞对,柳鹏程. 药物研发领域人工智能应用与创新发展策略探讨[J]. 中国新药与临床杂志, 2021, 40(6): 430-435.
Mao YD, Liu PC. Discussion on application and innovative development strategy of artificial intelligence in drug research and development field [J]. Chinese Journal of New Drugs and Clinical Remedies, 2021, 40(6): 430-435.

[10] Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces [J]. Phys Rev Lett, 2007, 98(14): 146401.

[11] Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking [J]. J Med Chem, 2006, 49(23): 6789-6801.

[12] Wu ZQ, Ramsundar B, Feinberg EN, et al. MoleculeNet: a benchmark for molecular machine learning [J]. Chem Sci, 2018, 9(2): 513-530.

[13] 颜彩琴,范睿琦,宁雨坪,等. 深度学习模型在中药毒性预警中的应用和前景[J]. 中国药理学与毒理学杂志, 2022, 36(3): 231-240.
Yan CQ, Fan RQ, Ning YP, et al. Applications and prospect of deep learning model in early warning of toxicity of traditional Chinese medicine [J]. Chinese Journal of Pharmacology and Toxicology, 2022, 36(3): 231-240.

[14] 宋学武,高慧儿,张戈. 基于人工智能的机器学习算法在个体化用药领域的应用进展[J]. 中国新药与临床杂志, 2021, 40(10): 683-688.
Song XW, Gao HE, Zhang Y. Advance in application of machine learning algorithm based on artificial intelligence in field of individualized medicine [J]. Chinese Journal of New Drugs and Clinical Remedies, 2021, 40(10): 683-688.

[15] 班娜,曹世奎. 中国创新药物研发现状[J]. 应用化学, 2022, 39(5): 857-858.
Ban N, Cao SK. Current situation of innovative drug research and development in China [J]. Chinese Journal of Applied Chemistry, 2022, 39(5): 857-858.

[16] 张莎,祝德秋. 基于美国FAERS数据库的新型降脂药前蛋白转化酶枯草杆菌素9抑制剂药物不良信号的数据挖掘与分析[J]. 中南药学, 2021, 19(12): 2582-2587.

- Zhang S, Zhu DQ. Adverse drug events signals of PCSK9 inhibitors based on the FAERS database[J]. *Central South Pharmacy*, 2021, 19(12): 2582-2587.
- [17] 王明阳. 基于人工智能技术的全新药物设计方法研究[D]. 杭州: 浙江大学, 2023.
- Wang MY. Methodological study on AI-based de novo drug design [D]. Hangzhou: Zhejiang University, 2023.
- [18] Ståhl N, Falkman G, Karlsson A, et al. Deep reinforcement learning for multiparameter optimization in de novo drug design[J]. *J Chem Inf Model*, 2019, 59(7): 3166-3176.
- [19] Russo DP, Zorn KM, Clark AM, et al. Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction[J]. *Mol Pharm*, 2018, 15(10): 4361-4370.
- [20] Segler MHS, 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.
- [21] Reker D, Rodrigues T, Schneider P, et al. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus[J]. *Proc Natl Acad Sci U S A*, 2014, 111(11): 4067-4072.
- [22] Olivecrona M, Blaschke T, Engkvist O, et al. Molecular de-novo design through deep reinforcement learning[J]. *J Cheminform*, 2017, 9(1): 48.
- [23] Turk S, Merget B, Rippmann F, et al. Coupling matched molecular pairs with machine learning for virtual compound optimization[J]. *J Chem Inf Model*, 2017, 57(12): 3079-3085.
- [24] Fraser K, Bruckner DM, Dordick JS. Advancing predictive hepatotoxicity at the intersection of experimental, in silico, and artificial intelligence technologies[J]. *Chem Res Toxicol*, 2018, 31(6): 412-430.
- [25] 毛火华, 裴奕钦. 药物发现中的人工智能[J]. *张江科技评论*, 2024(1): 6-9.
- Mao HH, Pei YQ. Artificial intelligence in drug discovery [J]. *Zhangjiang Technology Review*, 2024(1): 6-9.
- [26] Hughes TB, Swamidass SJ. Deep learning to predict the formation of quinone species in drug metabolism[J]. *Chem Res Toxicol*, 2017, 30(2): 642-656.
- [27] Lusci A, Pollastri G, Baldi P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules[J]. *J Chem Inf Model*, 2013, 53(7): 1563-1575.
- [28] Corey EJ, Long AK, Rubenstein SD. Computer-assisted analysis in organic synthesis[J]. *Science*, 1985, 228(4698): 408-418.
- [29] Schwaller P, Gaudin T, Lányi D, et al. "Found in translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models[J]. *Chem Sci*, 2018, 9(28): 6091-6098.
- [30] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI[J]. *Nature*, 2018, 555(7698): 604-610.

(编辑:黄开颜)