

# 人工智能在抗肿瘤药物研发中的应用与进展

管庆霞, 罗煜婷, 刘宇萌, 李莹, 杨寒, 李泌男, 李秀岩\* (黑龙江中医药大学, 哈尔滨 150040)

**摘要:** 随着人工智能的飞速发展以及人们对肿瘤疾病机制的认识逐渐加深, 人工智能在抗肿瘤药物的研发中起到了重要作用。本文简要介绍了人工智能的主要模型及在药物研发方面的优势, 并基于结构、配体、肿瘤细胞表型及电子健康记录四大方面综述了近年来人工智能在抗肿瘤药物研发中的应用与进展, 并总结了人工智能应用于抗肿瘤药物研发中的机遇与挑战, 以期从事相关领域的科研技术人员提供思路与参考。

**关键词:** 人工智能; 机器学习; 新药研发; 抗肿瘤

中图分类号: R979.1 文献标志码: A 文章编号: 1007-7693(2023)23-3318-06

DOI: 10.13748/j.cnki.issn1007-7693.20230857

引用本文: 管庆霞, 罗煜婷, 刘宇萌, 等. 人工智能在抗肿瘤药物研发中的应用与进展[J]. 中国现代应用药学, 2023, 40(23): 3318-3323.

## Application and Progress of Artificial Intelligence in Development of Anti-tumor Drugs

GUAN Qingxia, LUO Yuting, LIU Yumeng, LI Ying, YANG Han, LI Minan, LI Xiuyan\* (*Heilongjiang University of Traditional Chinese Medicine, Harbin 150040, China*)

**ABSTRACT:** With the rapid development of artificial intelligence and the gradual deepening of people's understanding of the mechanisms of tumor diseases, artificial intelligence has played an important role in the research and development of anti-tumor drugs. This article briefly introduces the main models of artificial intelligence and its advantages in drug research and development, and summarizes the application and progress of artificial intelligence in anti-tumor drug research and development in recent years from four aspects: structure based, ligand based, tumor cell phenotype, and electronic health records. It also summarizes the opportunities and challenges of artificial intelligence in anti-tumor drug research and development, hoping to provide ideas and references for scientific and technical personnel engaged in related fields.

**KEYWORDS:** artificial intelligence; machine learning; new drug development; anti-tumor

恶性肿瘤是威胁人类健康的主要因素之一<sup>[1]</sup>。治疗肿瘤的过程复杂且漫长, 现阶段治疗肿瘤的传统手段虽然有较好的疗效, 但不能完全有效控制肿瘤的复发及转移<sup>[2]</sup>。随着国际基因组计划的完成, 人类进入了基因组时代, 临床医学从宏观到微观的发展, 从细胞到分子水平的研究, 使得肿瘤学领域的精准肿瘤治疗飞速发展, 通过全基因组测序构建的肿瘤基因组推动了肿瘤分子发生机制的研究, 以此发现新的抗肿瘤靶标, 全面革新了肿瘤治疗模式, 将抗肿瘤药物推向一个新时代。然而新型抗肿瘤药物的发现是一个缓慢长期的过程, 通常需要 10~20 年的时间, 花费的成本从 5 亿美元到数 10 亿美元不等, 具有费用高、周期长、成功率低等特点<sup>[3-4]</sup>。随着计算机技术的发展, 许多科研人员选择用虚拟筛选技术辅助抗肿瘤药物的早期开发, 计算机技术成为寻找潜在抗肿瘤药

物成本更低, 更有效率的方法。

随着大数据时代的到来及计算机数据处理能力的快速提高, 人工智能(artificial intelligence, AI)逐渐成为缩短药物研发周期和降低成本的先进工具, 将 AI 应用于抗肿瘤药物的虚拟筛选中, 可为药物如何与靶标结合以及它们的理化特性如何与表型变化相关提供了新的见解。目前通过 AI 技术已找到多种依赖性癌基因<sup>[5]</sup>, 并且能通过 AI 技术筛选大量关于肿瘤细胞和小分子数据, 以此寻找小分子药物与依赖性癌基因之间的关联。例如蛋白激酶在细胞生长、分化和凋亡中起着至关重要的作用<sup>[6]</sup>, 蛋白激酶功能异常可导致许多严重的疾病, 如肺癌<sup>[7-8]</sup>、卵巢癌<sup>[9]</sup>、乳腺癌<sup>[10-11]</sup>、前列腺癌<sup>[12-13]</sup>等, 其中蛋白激酶抑制剂(protein kinase inhibitor, PKI)为抗肿瘤药物研究的重要领域之一<sup>[14]</sup>, 而 AI 在筛选蛋白激酶特异性小分子抑制剂方面有着明显优

**基金项目:** 黑龙江自然科学基金联合引导项目(LH2022H085); 黑龙江省中医药科研项目(ZHY2022-111); 黑龙江省自然科学基金面上项目(H2016076)

**作者简介:** 管庆霞, 女, 博士, 副教授 E-mail: gqxwyb@163.com

**\*通信作者:** 李秀岩, 女, 博士, 讲师 E-mail: 812132357@qq.com

势。将 AI 技术引入挖掘分子对肿瘤治疗靶标的活性和肿瘤细胞系敏感性的研究,是研发新型肿瘤药物的发展方向。本文综述了近年来 AI 在抗肿瘤药物研发中的应用与进展,以期从事相关领域的科研技术人员提供参考。

## 1 AI 算法模型简要概况

一般来说, AI 是指机器执行相关任务以响应一系列环境的能力:如为了预测与药物发现相关的结果,机器需要某些算法来处理现有数据并识别功能属性的模式,此过程被视为机器学习。药物研发中最常见的机器学习算法有 k 近邻(k-nearest neighbours, kNN)、朴素贝叶斯(Naïve Bayesian classifier, NB)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)和人工神经网络(artificial neural network, ANN)等。kNN 指在特征空间中,如果一个样本附近的 k 个最近(即特征空间中最近)样本的大多数属于某一个类别,则该样本也属于这个类别<sup>[15]</sup>。NB 是贝叶斯分类器中的一种模型,可用已知类别的数据集训练模型,从而实现对未知类别数据的类别判断<sup>[16]</sup>。RF 是在原始数据集中随机抽样,构成许多不同数据集,再根据数据集搭建不同的决策树模型,最后根据这些决策树模型的平均值(针对回归模型)或者投票情况(针对分类模型)来获取最终结果。SVM 是一种二类分类模型,其基本模型定义为特征空间上的间隔最大的线性分类器,其学习策略便是间隔最大化,最终可转化为一个凸二次规划问题的求解<sup>[17]</sup>。ANN 是以数学模型模拟神经元活动,基于模仿大脑神经网络结构和功能而建立的一种信息处理系统<sup>[18]</sup>。这些算法在抗肿瘤药物预测方面各有其优势,见表 1。

表 1 AI 算法模型在抗肿瘤药物预测方面的主要优势

Tab.1 Major advantages of AI algorithm models in antitumor drug prediction

模型	应用优势	参考文献
朴素贝叶斯	适用于虚拟筛选分类和获取特异性结合于癌症靶点的分子骨架	[19]
K 近邻	对于预测多靶点结合活性等多任务学习具有明显优势	[20]
支持向量机	可用于抗肿瘤化合物分类和预测其属性	[21]
随机森林	改善定量构效关系数据预测,也可用于对接打分函数以及预测靶标-配体结合亲和力研究	[22]
人工神经网络	常应用于潜在抗癌药物靶标识别、分类、定量构效关系以及靶标-配体结合亲和力等研究	[23]

## 2 基于结构的机器学习

一般基于结构的虚拟筛选(structure-based

virtual screening, SBVS)计算模型旨在优先考虑可靶向目标靶蛋白的小分子,典型的方法是预测小分子和目标靶蛋白之间的结合亲和力(或相关亲和力评分),然后根据亲和力大小对化合物进行优先级排序,见图 1。

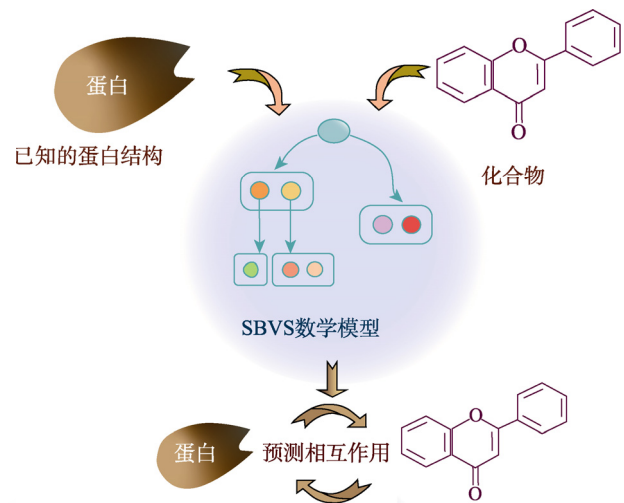


图 1 SBVS 在药物研发中的应用

Fig.1 Application of SBVS in drug development

### 2.1 预测药物-肿瘤靶标亲和力

加速 PKI 发现的第一步就是快速识别小分子和激酶是否相互作用。在实验上,这些相互作用可以通过体外结合试验确定,而体外结合试验是一种昂贵且费力的方法,无法大规模应用。因此,迫切需要开发统计学上有效的方法来识别 PKI 之间的相互作用。蛋白质突变是癌症发生的主要原因,杨子怡等<sup>[24]</sup>提出了一种机器学习方法,称为 SPLD 额外树(SPLD Extra Trees),该方法可以准确预测癌蛋白突变时与配体结合亲和力的变化,识别引起耐药性的突变基因,并证实了用此方法预测 PKI 耐药性的能力。Durán 等<sup>[25]</sup>运用 ANN 和采样方法研究了卡博替尼与程序性细胞死亡蛋白-1(programmed death 1, PD-1)抑制剂在转移性肾透明细胞癌中的潜在协同作用的机制,通过构建 ANN 证明了 PD-1 抑制剂可以通过调节血管生成和 T 细胞毒性来增强卡博替尼的抗血管生成和促进凋亡作用,并且预测单独使用卡博替尼可恢复细胞黏附并阻碍肿瘤增殖和侵袭。Cao 等<sup>[26]</sup>首次将激酶-抑制剂之间的定量结合亲和力作为定义抑制剂与激酶相互作用的指标,然后利用化学基因组学框架,使用一组无偏倚的综合特征集(药物描述符和蛋白描述符)和 RF 构建预测模型来准确分类激酶-抑制剂,推动了抗肿瘤药物的发现。

## 2.2 分子对接算法的改善

当前配体与靶点蛋白之间的结合姿态以及结合亲和力的强度主要通过分子对接预测,再通过打分函数对该姿态和强度进行结合自由能计算并排序。现有的经典评分函数很多,但是对接不准确且误报率较高<sup>[27]</sup>,因此利用机器学习等新方法对于提高对接精度至关重要。利用机器学习,评分函数可以向无偏见的方式大幅改进,因为药物-靶点关系可直接从训练数据中推断且不受推定关系的影响,从而更好地理解驱动药物-靶点相互作用的力量。Wan 等<sup>[28]</sup>建立 SVM 模型用于分子对接和药效团的虚拟筛选,从数据库中发现新的酪氨酸 PKI,大大提高了 PKI 的筛选效率。Li 等<sup>[29]</sup>用 SVM 模型研究出来了一种靶标特异性打分函数,用此函数针对激酶表皮生长因子受体筛选其抑制剂,成功发现了数个活性分子。郑良珍等<sup>[30]</sup>将数据驱动模型和经验评分函数相结合,使用基于 RF 模型的校正项增强了传统评分函数,以此开发出可靠的蛋白质配体评分函数,大大增强了评分函数基于不同基准对接和筛选任务的预测能力。哺乳动物靶点雷帕霉素 PKI 的发现一直是抗肿瘤药物的研究热点,李东东等<sup>[31]</sup>基于偏最小二乘法 (partial least square, PLS)对获得的评分结果进行共识评分模型训练,并采用留一法验证并用于新型雷帕霉素 PKI 的对接虚拟筛选。秦魏等<sup>[32]</sup>采用 SYBYL 软件对自主构建化合物库进行虚拟筛选,为靶向聚 ADP 核糖聚合酶-14 蛋白催化结构域抗肿瘤药物研发提供潜在先导化合物,并能通过此软件评估每一次分子对接的结果。

### 3 基于配体的机器学习

阐明药物-靶点相互作用的第二种方法是基于配体的化学信息学方法。基于配体的虚拟筛选 (ligand-based virtual screening, LBVS)是在已知的配体的基础上进行药物筛选与设计<sup>[33]</sup>,通过对一系列具有同样母核结构的化合物进行分析,得到其构效关系并对小分子进行修饰和优化,见图 2。

#### 3.1 预测小分子抑制剂的生物学活性

Nijjima 等<sup>[34]</sup>运用了一种分解激酶谱数据的反卷积方法,可以从大规模谱数据中获得关于抑制剂交叉反应性的信息,这种方法不仅可以在激酶组范围内预测给定化合物的活性,还可以提取与活性相关的残基片段对。Lapins 等<sup>[35]</sup>运用 kNN 算法,应用蛋白化学计量学模型将 317 种野生型和

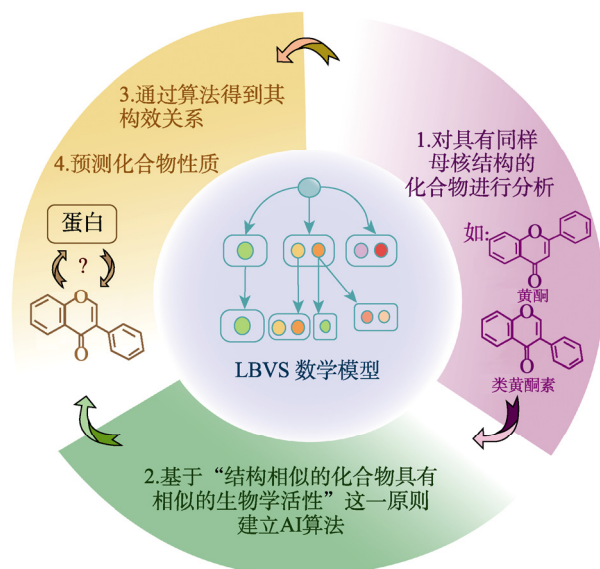


图 2 LBVS 在药物研发中的应用

Fig. 2 Application of LBVS in drug development

突变型激酶以及 38 种抑制剂(12 046 种抑制剂-激酶组合)的特性与各自组合的相互作用解离常数关联起来,比较了描述蛋白激酶的 6 种方法以及几种线性和非线性相关方法,将最好的模型用 SVM 或 PLS 对潜在结构进行相关性分析,以此预测小分子抑制剂的生物学活性。Merget 等<sup>[36]</sup>运用 RF 算法报告了基于配体的 280 多种激酶活性预测模型的生成。张建华<sup>[37]</sup>通过收集了 147 个激酶靶标和抑制剂筛选活性信息,采用 NB 方法,构建了基于 882 个基于原子中心片段的 PKI 活性预测模型。王领<sup>[38]</sup>基于 NCI-60 细胞系的肿瘤细胞筛选大数据,构建完善的抗肿瘤细胞活性化合物数据库、数据标注及数据标准化。分别采用 NB、RF、单向树机器学习方法,以原子中心片段和分子指纹构建了细胞水平的抗肿瘤化合物活性预测模型。

#### 3.2 预测小分子抑制剂的吸收、分布、代谢、排泄和毒性

由于药物研发难度较大、过程复杂,这就要求研发人员掌握候选化合物的结构、安全性、药动学性质等特征参数,即药物的吸收、分布、代谢、排泄和毒性(absorption distribution metabolism excretion and toxicity, ADMET)<sup>[39]</sup>。秦雅琴等<sup>[40]</sup>为提高乳腺癌药物虚拟筛选中药物 ADMET 的预测能力,提出了一种基于逻辑回归、NB、梯度提升决策树 (gradient boosting decision tree, GBDT)以及贝叶斯超参数优化算法 (Bayesian hyperparametric

optimization, BHO)等算法设计的抗乳腺癌药物定量结构-ADMET 性质预测模型。许美贤等<sup>[41]</sup>为加快抗乳腺癌候选药物的研发进程,应用粒子群优化神经网络算法以及基于粒子群优化支持向量机来预测化合物的 ADMET 性质,从而寻找到 ADMET 性质较好,活性较高的抗乳腺癌化合物分子。

### 3.3 预测小分子抑制剂靶标

现在许多小分子之间的关联已使用化学相似性技术阐明,但在某些情况下微小的化学修饰(即微小的官能团变化)会导致小分子活性的巨大变化,也称为“活性悬崖”,从而导致假阳性关联<sup>[42]</sup>。为了规避上述限制,可将机器学习纳入药物特征与生物活性关联性的考察,这通常被称为药物定量构效关系(quantitative structure-activity relationship, QSAR)。Cui 等<sup>[43]</sup>提出了基于文化差分进化算法的最小二乘支持向量机(cultural differential evolution algorithm least square support vector machine, CDE-LSSVM)模型,运用此模型建立蛋白酶抑制剂的 QSAR 具有很好的拟合精度与预测精度,且优于最小二乘支持向量机、反向传播神经网络和径向基神经网络。沈泓佑等<sup>[44]</sup>以深度学习算法作为建模工具,分别建立了葱醌化合物结构特征与抗肿瘤活性、正常细胞毒性之间初步的 QSAR 模型。蛋白酶活化受体-4(protease activated receptors 4, PAR-4)对肠癌细胞增殖迁移有较大影响,马威<sup>[45]</sup>以机器学习模型为基础,以“ChemDiv”化合物库为分子对接化合物数据来源,得到训练数据集后运用机器学习模型进行数据训练得到虚拟筛选模型,对 PAR-4 进行三维结构的建模及模型的优化与验证。

### 4 基于肿瘤细胞表型数据的机器学习

细胞表型是涉及基因蛋白表达的多个细胞过程的集合体,这些过程导致细胞特定的形态和功能,肿瘤的细胞表型和正常细胞表型之间有明显区分。尽管目前已成功将靶点的抑制与所需的临床结果相联系,但对药物与靶点之间完整的相互作用过程认知情况还很有限。而通过剂量反应、细胞死亡和细胞凋亡分析研究药物对细胞和复杂组织的直接表型效应,可以部分解决这一问题。

因此通过建立大规模数据集,并利用 AI 预测抗肿瘤药物-肿瘤细胞表型的关联是一种抗肿瘤药物发现的可靠方法,见图 3。

Kadurin 等<sup>[46]</sup>将对抗性自动编码器应用于 NCI-60 细胞系测定的全剂量反应数据以此开发深

度学习模型,通过算法输送>7 200 万种化合物的 PubChem 数据库,并能够通过寻找已经证明具有抗肿瘤活性的分子以及揭示具有潜在抗肿瘤活性的新分子来验证他们的模型。此外高通量成像分析也是抗肿瘤药物筛选中表型数据的丰富来源,例如暴露于药物的细胞的形态变化可以通过细胞染色显微镜捕获,然后再使用软件提取其形态信息,并转换为基于图像的化合物指纹图谱,适合机器学习以预测化合物活性<sup>[47]</sup>。在最近的一项研究中,Simm 等<sup>[48]</sup>在糖皮质激素受体易位的三通道筛选中为化合物处理的细胞创建了一个基于高维图像的指纹,针对脑神经胶质瘤细胞筛选出 524 371 种化合物,并提取每种化合物的特征形成指纹,将这些指纹图谱与来自正交测定的已知药物生物活性数据相结合,以此预测出了其他对脑神经胶质瘤细胞有特异性活性的小分子。



图 3 肿瘤细胞表型数据在药物研发中的应用

Fig. 3 Application of tumor cell phenotype data in drug development

### 5 基于电子健康记录(electronic health records, EHR)的机器学习

除了利用化学信息学、表型和转录组学数据集外,患者的 EHR 还包含大量信息,可用于抗肿瘤药物发现,再利用和药物不良事件预测。自然语言处理的最新进展使研究人员能够获取大部分非结构化数据源(即临床笔记,网络论坛帖子)并将其转换为机器可解释的格式,从而实现了机器学习分析的应用<sup>[49]</sup>,见图 4。

在一项基于 EHR 的研究中,Xu 等<sup>[50]</sup>开发了一个机器学习框架,将梅奥诊所和范德比尔特大学医学中心的 2 个 EHR 数据库整合,该课题组在 1995—2010 年记录了 2 个独立的队列,总计>100 000 例患者,并在控制其他变量(如吸烟)的同时,发现了抗糖尿病药物二甲双胍与提高肿瘤存活率之间的关联,二甲双胍作为抗肿瘤疗法的再利用潜力也在乳腺癌的几项临床前和流行病学研究中得到证实。



图4 电子健康记录在药物研发中的应用

Fig. 4 Application of electronic health records in drug development

## 6 AI 在肿瘤药物研发中的挑战

目前, AI 应用在药物筛选方面的主要限制是可用肿瘤数据库的质量和数量。随着高度自动化高通量筛选平台的成本不断降低, 药物检测筛选产生的数据量呈指数级增长, 但是并不能保证数据的精确度, 而且有关癌症的大数据仍然有限。所以机器学习受到所执行的样本数量和可用数据源的限制, 现阶段用于抗肿瘤药物筛选过程的临床前方法虽然省时省力但是并不完美。且目前机器学习技术由于缺乏可解释性和对单一疗法的关注, 使得大多数预测到的药物尚未达到临床实践。各种算法模型中复杂和非线性的输入端与其相应输出端之间的连接复杂且不易理解, 对于评估和理解 AI 的决策也是一项艰巨任务。

## 7 讨论

随着癌症基因组图谱等大规模数据收集工作的展开, 为 AI 在药物发现应用方面的条件奠定了基础, AI 在医药领域的应用也成了抗肿瘤药物筛选的新方式。但是 AI 方法在现阶段仍然是一个“黑匣子”, 例如在数据训练阶段, 神经网络只被允许特定输入, 但是输入的特征没有明确指定, 甚至网络的创建者也可能不知道中间阶段正在检查什么, 或者为什么模型会得出此结论。尽管大数据有助于在肿瘤治疗中发现可能具有潜力的候选药物, 但最终还是得使用实验方法测试这些候选药物, 以验证其功效和毒性并研究作用机制。所以需要注意的是, 尽管 AI 的发展十分迅猛, 但它的出现只能助力寻找新抗肿瘤小分子从而促使工作效率的提高, 研究员不能过度依赖于此。在合理利用 AI 的基础上, 将 AI 技术引入抗肿瘤药物及其作用机制的研究, 用于药物设计过程, 不但减少了临床前试验相关的健康危害, 也大大降低了成本。深入挖掘药物

对肿瘤治疗靶标的敏感性是高效新型肿瘤药物开发的发展方向。

## REFERENCES

- [1] SONG Z D, WANG M J, GE Y, et al. Tyrosine phosphatase SHP<sub>2</sub> inhibitors in tumor-targeted therapies[J]. *Acta Pharm Sin B*, 2021, 11(1): 13-29.
- [2] PEER D, KARP J M, HONG S, et al. Nanocarriers as an emerging platform for cancer therapy[J]. *Nat Nanotechnol*, 2007, 2(12): 751-760.
- [3] LIU B Q, HE H Q, LUO H Y, et al. Artificial intelligence and big data facilitated targeted drug discovery[J]. *Stroke Vasc Neurol*, 2019, 4(4): 206-213.
- [4] MOORE T J, ZHANG H Z, ANDERSON G, et al. Estimated costs of pivotal trials for novel therapeutic agents approved by the US food and drug administration, 2015-2016[J]. *JAMA Intern Med*, 2018, 178(11): 1451-1457.
- [5] WEINSTEIN I B. Cancer. Addiction to oncogenes: The Achilles heal of cancer[J]. *Science*, 2002, 297(5578): 63-64.
- [6] ZHOU X J, KUANG Y C, LIANG S M, et al. Metformin inhibits cell proliferation in SKM-1 cells via AMPK-mediated cell cycle arrest[J]. *J Pharmacol Sci*, 2019, 141(4): 146-152.
- [7] PU X, WANG L, CHANG J Y, et al. Inflammation-related genetic variants predict toxicity following definitive radiotherapy for lung cancer[J]. *Clin Pharmacol Ther*, 2014, 96(5): 609-615.
- [8] ZHANG N, SU Y S, XU L J. Targeting PKC $\epsilon$  by miR-143 regulates cell apoptosis in lung cancer[J]. *FEBS Lett*, 2013, 587(22): 3661-3667.
- [9] CUI Y Y, YAN L, ZHOU J, et al. The role of peptidylarginine deiminase 4 in ovarian cancer cell tumorigenesis and invasion[J]. *Tumour Biol*, 2016, 37(4): 5375-5383.
- [10] SHAO W F, LI F X, TANG R, et al. Nanog promotes the invasion of breast cancer cells by increasing PKC $\epsilon$  expression[J]. *J South Med Univ*, 2016, 36(5): 639-644.
- [11] CHAUVIN L, GOUPILLE C, BLANC C, et al. Long chain n-3 polyunsaturated fatty acids increase the efficacy of docetaxel in mammary cancer cells by downregulating Akt and PKC $\epsilon$ / $\delta$ -induced ERK pathways[J]. *Biochim Biophys Acta*, 2016, 1861(4): 380-390.
- [12] HAFEEZ B B, MESKE L, SINGH A, et al. Tissue-specific conditional PKC $\epsilon$  knockout mice: A model to precisely reveal PKC $\epsilon$  functional role in initiation, promotion and progression of cancer[J]. *Oncotarget*, 2016, 7(22): 33069-33080.
- [13] GARG R, BLANDO J M, PEREZ C J, et al. Protein kinase C *Epsilon* cooperates with PTEN loss for prostate tumorigenesis through the CXCL13-CXCR5 pathway[J]. *Cell Rep*, 2017, 19(2): 375-388.
- [14] DU J. Computer aided molecular design and simulation of kinase inhibitors[D]. Lanzhou: Lanzhou University, 2011.
- [15] COVER T, HART P. Nearest neighbor pattern classification[J]. *IEEE Trans Inf Theory*, 1967, 13(1): 21-27.
- [16] LÓPEZ PUGA J, KRZYWINSKI M, ALTMAN N. Points of significance: Bayes' theorem[J]. *Nat Methods*, 2015, 12(4): 277-278.
- [17] CORTES C, VAPNIK V. Support-vector networks[J]. *Mach Lang*, 1995, 20(3): 273-297.
- [18] ROSENBLATT F. The perceptron: A probabilistic model for

- information storage and organization in the brain[J]. *Psychol Rev*, 1958, 65(6): 386-408.
- [19] BENDER A. Bayesian methods in virtual screening and chemical biology[J]. *Methods Mol Biol*, 2011(672): 175-196.
- [20] CARPENTER K A, HUANG X D. Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: A review[J]. *Curr Pharm Des*, 2018, 24(28): 3347-3358.
- [21] SHI Z, MA X H, QIN C, et al. Combinatorial support vector machines approach for virtual screening of selective multi-target serotonin reuptake inhibitors from large compound libraries[J]. *J Mol Graph Model*, 2012(32): 49-66.
- [22] LAVECCHIA A. Machine-learning approaches in drug discovery: Methods and applications[J]. *Drug Discov Today*, 2015, 20(3): 318-331.
- [23] ZHONG F S, XING J, LI X T, et al. Artificial intelligence in drug design[J]. *Sci China Life Sci*, 2018, 61(10): 1191-1204.
- [24] YANG Z Y, YE Z F, XIAO Y J, et al. SPLDExtraTrees: Robust machine learning approach for predicting kinase inhibitor resistance[J]. *Brief Bioinform*, 2022, 23(3): bbac050.
- [25] DURÁN I, CASTELLANO D, PUENTE J, et al. Exploring the synergistic effects of cabozantinib and a programmed cell death protein 1 inhibitor in metastatic renal cell carcinoma with machine learning[J]. *Oncotarget*, 2022(13): 237-256.
- [26] CAO D S, ZHOU G H, LIU S, et al. Large-scale prediction of human kinase-inhibitor interactions using protein sequences and molecular topological structures[J]. *Anal Chim Acta*, 2013(792): 10-18.
- [27] CHEN Y C. Beware of docking![J]. *Trends Pharmacol Sci*, 2015, 36(2): 78-95.
- [28] WAN H L, WANG Z R, LI L L, et al. Discovery of novel Bruton's tyrosine kinase inhibitors using a hybrid protocol of virtual screening approaches based on SVM model, pharmacophore and molecular docking[J]. *Chem Biol Drug Des*, 2012, 80(3): 366-373.
- [29] LI L W, KHANNA M, JO I, et al. Target-specific support vector machine scoring in structure-based virtual screening: Computational validation, *in vitro* testing in kinases, and effects on lung cancer cell proliferation[J]. *J Chem Inf Model*, 2011, 51(4): 755-759.
- [30] ZHENG L Z, MENG J T, JIANG K, et al. Improving protein-ligand docking and screening accuracies by incorporating a scoring function correction term[J]. *Brief Bioinform*, 2022, 23(3): bbac051.
- [31] LI D D, MENG X F, WANG Q, et al. Consensus scoring model for the molecular docking study of mTOR kinase inhibitor[J]. *J Mol Graph Model*, 2018(79): 81-87.
- [32] QIN W, LI C, WANG R, et al. Virtual screening of PARP14 catalytic domain inhibitors[J]. *Clin Res Pract(临床医学研究与实践)*, 2020, 5(30): 15-17.
- [33] JOHNSON M A, MAGGIORA G M. Concepts and applications of molecular similarity[M]. New York: Wiley, 1990.
- [34] NIJIMA S, SHIRAIISHI A, OKUNO Y. Dissecting kinase profiling data to predict activity and understand cross-reactivity of kinase inhibitors[J]. *J Chem Inf Model*, 2012, 52(4): 901-912.
- [35] LAPINS M, WIKBERG J E. Kinome-wide interaction modelling using alignment-based and alignment-independent approaches for kinase description and linear and non-linear data analysis techniques[J]. *BMC Bioinformatics*, 2010(11): 339.
- [36] MERGET B, TURK S, EID S, et al. Profiling prediction of kinase inhibitors: Toward the virtual assay[J]. *J Med Chem*, 2017, 60(1): 474-485.
- [37] ZHANG J H. Study on activity and selectivity prediction of kinase inhibitors[D]. Guangzhou: South China University of Technology, 2021.
- [38] 王领. 基于细胞-靶标组合活性预测的抗癌靶向 mTOR 抑制剂发现与作用机制研究[D]. 广州: 华南理工大学, 2019.
- [39] JIANG D J, LEI T L, WANG Z, et al. ADMET evaluation in drug discovery. 20. Prediction of breast cancer resistance protein inhibition through machine learning[J]. *J Cheminform*, 2020, 12(1): 16.
- [40] QIN Y Q, XIA Y L, LU M Y, et al. ADMET property prediction model of anti-breast cancer active compounds[J]. *J Yunnan Univ(云南大学学报: 自然科学版)*, 2022, 44(6): 1127-1134.
- [41] XU M X, ZHENG Y, LI Y J, et al. Prediction of properties of anti-breast cancer drugs based on PSO-BP neural network and PSO-SVM[J]. *J Nanjing Univ Inf Sci Technol(南京信息工程大学学报: 自然科学版)*, 2023, 15(1): 51-65.
- [42] HU Y, STUMPFE D, BAJORATH J. Advancing the activity cliff concept[J]. *F1000Res*, 2013(2): 199.
- [43] CUI W T, HU C P, YAN X F. Least Square support vector machine based on cultural differential evolution algorithm and its application to QSAR modeling[J]. *J East China Univ Sci Technol(华东理工大学学报: 自然科学版)*, 2010, 36(1): 121-125.
- [44] SHEN H Y, HE X N, LU S F, et al. Deep learning design and synthesis of novel anthraquinone topoisomerase II inhibitors[J]. *Chin J Synth Chem(合成化学)*, 2021, 29(4): 261-267.
- [45] MA W. Construction of virtual screening model based on machine learning and homologous modeling and structural verification of PAR4 protein[D]. Shanghai: East China University of Science and Technology, 2021.
- [46] KADURIN A, ALIPER A, KAZENNOV A, et al. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology[J]. *Oncotarget*, 2017, 8(7): 10883-10890.
- [47] SCHEEDER C, HEIGWER F, BOUTROS M. Machine learning and image-based profiling in drug discovery[J]. *Curr Opin Syst Biol*, 2018(10): 43-52.
- [48] 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.e3.
- [49] LI F, LIU W S, YU H. Extraction of information related to adverse drug events from electronic health record notes: Design of an end-to-end model based on deep learning[J]. *JMIR Med Inform*, 2018, 6(4): e12159.
- [50] XU H, ALDRICH M C, CHEN Q X, 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(1): 179-191.

收稿日期: 2023-03-31  
 (本文责编: 李艳芳)