

基于人工智能的药物筛选算法优化与在新药研发中的应用研究

郑春新

台州黄罐健源食品有限公司, 浙江 台州 310000

摘要: 人工智能技术在新药研发领域的应用日益广泛, 特别是在药物筛选环节展现出强大潜力。传统药物筛选方法耗时长、成本高, 难以满足现代药物研发的效率需求。本文基于人工智能的药物筛选算法优化为切入点, 系统分析了常见算法在分子特征提取、虚拟筛选和药效预测中的应用现状与挑战, 重点探讨算法优化策略及其对提升筛选效率和准确性的影响。通过案例研究, 展示优化算法在新药研发中的具体应用效果, 旨在为药物研发提供科学、有效的技术支持, 推动新药发现的智能化进程。

关键词: 人工智能; 药物筛选; 算法优化; 新药研发; 虚拟筛选

DOI: 10.63887/jmph.2025.1.2.15

引言

新药研发是现代医学和生命科学领域最具挑战性的任务之一, 涉及复杂的分子设计与验证过程。传统药物筛选方法通常依赖于高通量实验筛选, 虽然技术不断进步, 但仍存在时间长、成本高、成功率低等瓶颈^[1]。近年来, 随着人工智能技术的快速发展, 尤其是深度学习和机器学习算法在生物信息学和化学信息学领域的融合, 为药物筛选提供了新的思路 and 工具^[2]。人工智能能够利用大量分子数据, 挖掘复杂的结构与活性关系, 实现高效的虚拟筛选和药效预测, 极大地提升了新药研发的效率和成功率。

然而, 人工智能算法在药物筛选中的应用也面临多方面挑战, 包括数据质量参差不齐、特征提取不足、模型过拟合以及解释性差等问题^[4]。如何针对这些问题优化算法, 提升模型的泛化能力和预测准确度, 成为当前研究的热点。本文将围绕人工智能药物筛选算法的优化展开, 系统探讨其在新药研发各阶段的应用, 结合具体案例分析算法改进效果, 旨在为科研人员和药企提供参考, 促进新药研发的智能化

升级。

1 人工智能药物筛选算法的现状与挑战

1.1 药物筛选中的人工智能技术应用现状

药物筛选作为新药研发的核心环节, 长期以来面临着实验周期长、成本高昂和成功率低的问题。传统的高通量筛选虽提高了效率, 但依然无法满足现代药物研发对速度和精度的双重需求。近年来, 人工智能技术的兴起为药物筛选带来了革命性的变革。机器学习和深度学习, 尤其是支持向量机 (SVM)、随机森林 (RF) 以及卷积神经网络 (CNN) 等, 成为了处理海量分子数据和预测药物活性的主要工具。通过从大量已知药物分子及其生物活性数据中学习, 这些算法能够快速识别潜在有效化合物, 从而大幅缩短药物发现周期并降低实验成本^[5]。

其中, 基于图神经网络 (GNN) 的模型因其能够直接处理分子结构图, 捕获分子间复杂的空间和化学关系, 表现出优异的预测能力。

GNN 通过模拟分子中原子及其键的连接关系,实现对分子特征的深入挖掘,极大提高了虚拟筛选的准确率。配合丰富的化学数据库资源,人工智能驱动的虚拟筛选已被众多制药企业和科研机构广泛采用,有效支持了新药设计的创新。此外,人工智能还被用于药物-靶标相互作用预测、ADMET(吸收、分布、代谢、排泄、毒性)性质评估等多个环节,形成了覆盖药物研发全流程的智能化体系。

1.2 药物筛选算法面临的主要挑战

尽管人工智能技术带来了显著进步,药物筛选领域的算法优化仍面临多重挑战。首先,药物筛选数据本身存在不均衡和噪声问题。实验数据往往带有测量误差,且有效活性分子远少于无效分子,导致模型在训练时容易偏向多数类,影响预测的准确性和泛化能力。其次,如何从分子数据中提取有效特征仍是难点。分子结构复杂多样,单一的指纹描述或简单特征难以全面捕捉其化学和生物活性信息,直接限制了模型性能。

此外,过拟合现象普遍存在。由于新药数据集有限,模型在训练时容易记忆训练样本的噪声和细节,导致对新样本预测失效。缺乏足够的解释性也降低了 AI 模型在药物研发中的信任度,科研人员和监管机构往往要求对预测结果有明确的科学依据,而多数深度学习模型呈“黑盒”状态,难以解释其决策过程。因此,针对这些问题,如何设计更加稳健、高效且具解释性的算法,成为推动人工智能药物筛选技术发展的关键。

2 药物筛选算法的优化策略

2.1 数据预处理与特征工程的优化

数据是人工智能药物筛选的基础,而高质量的数据预处理对提升模型性能至关重要。药物筛选数据往往存在缺失值、异常点以及类别不平衡等问题,这些都会影响算法的学习效果。优化数据预处理流程,首先需要数据进

行清洗,剔除噪声数据和无效样本,保证输入数据的准确性和一致性。其次,针对类别不平衡问题,可采用过采样、欠采样或合成少数类样本等技术,以平衡活性分子与非活性分子的比例,避免模型偏向多数类。此外,采用多源数据融合技术,将结构数据、生物活性数据以及实验条件等多维信息融合,能够更全面地描述药物分子特征,增强模型的学习能力。

特征工程作为连接数据与模型的关键环节,也在优化药物筛选算法中发挥着重要作用。传统的分子指纹和描述符虽然广泛应用,但其表达能力有限,难以捕捉分子内部的复杂信息。近年来,基于深度学习的自动特征提取逐渐兴起,如图神经网络(GNN)通过直接学习分子图结构,自动挖掘分子间隐含的特征,有效提升了筛选准确率。除此之外,利用迁移学习和多任务学习技术,可以借助相关任务的知识,提升模型对新药物结构的泛化能力,减轻因数据不足导致的过拟合问题。综合利用多种特征优化技术,能够显著提升药物筛选模型的效果和稳定性。

2.2 模型结构与训练方法的改进

优化模型结构和训练方法是提升药物筛选算法性能的另一重要方向。当前深度学习模型结构多样,包括卷积神经网络(CNN)、循环神经网络(RNN)和图神经网络(GNN)等,不同结构适应不同类型的药物分子数据。针对药物筛选的特殊需求,设计适合分子数据特点的模型结构非常关键。例如,GNN 因能有效处理分子图结构,近年来被广泛应用于药物活性预测和虚拟筛选,相关模型在多个公开数据集上取得领先成绩。此外,混合模型通过结合多种神经网络结构,进一步提升了模型的表达能力和鲁棒性。

训练方法的改进主要集中在防止过拟合、提升泛化能力以及提高训练效率上。正则化技术如 Dropout、L2 正则化被广泛应用于深度学习模型,以避免模型过拟合训练数据。采用早

停法监控验证集性能,及时调整训练过程,进一步保证模型的泛化能力。此外,优化算法如 Adam、RMSProp 的应用,提升了模型训练的收敛速度和稳定性。近年来,联邦学习和自监督学习等新兴训练方法开始应用于药物筛选领域,利用分布式数据资源和无标签数据,进一步扩展了模型训练的边界和深度。这些模型和训练方法的不断优化,推动了人工智能药物筛选技术向更高效、更精准的方向发展。

3 人工智能药物筛选算法在新药研发中的应用

3.1 虚拟筛选中的应用实践

虚拟筛选是新药研发中利用计算技术快速筛选潜在活性分子的关键步骤。传统虚拟筛选主要依赖分子对接和基于结构的评价方法,效率和准确性受限。人工智能技术尤其是深度学习的引入,为虚拟筛选带来了显著提升。通过训练大量分子-靶标相互作用数据,AI 模型能够自动识别影响分子活性的关键特征,实现高效的分子筛选。

具体应用中,基于图神经网络的模型能够精准捕捉分子的拓扑结构和化学信息,识别出具备高亲和力的候选化合物。结合药物靶点的结构信息,AI 模型不仅提升了虚拟筛选的准确率,还大幅缩短了筛选时间。例如,某些制药公司通过自主开发的 AI 平台,成功筛选出数个临床前候选药物,极大地加速了研发进度。此外,虚拟筛选与实验验证相结合的模式,也逐渐成为新药发现的主流方法,提高了整体研发的成功率和成本效益。这种协同模式不仅节省了大量实验资源,还能通过持续反馈优化算法,推动筛选技术不断进步。

3.2 药效预测及安全性评估

新药研发不仅需要筛选出具备活性的候选分子,更重要的是准确预测药物的疗效和安全性。人工智能技术通过大数据分析和机器学习模型,有效提升了药效预测和毒性评估的准

确性。药效预测主要涉及药物-靶标相互作用、药代动力学及药效动力学的模拟,AI 模型能够整合多源数据,对候选药物的生物活性进行定量评估。

在安全性评估方面,传统方法通常依赖动物实验和临床试验,耗时且成本高。人工智能通过预测化合物的 ADMET 特性(吸收、分布、代谢、排泄及毒性),提前筛查出潜在的安全风险,显著降低了研发风险。深度学习模型通过分析化学结构与毒性数据的复杂关系,实现对毒副作用的早期识别,为新药的临床试验设计提供科学依据。这些 AI 驱动的药效与安全性预测技术,正在逐步改变传统新药研发流程,提高研发效率和安全保障水平。同时,AI 辅助的风险预警系统能动态监控药物开发过程中的不良反应,促进更安全、更高效的药物研发实践。

4 未来发展趋势与挑战

4.1 人工智能药物筛选技术的发展方向

随着计算能力和数据资源的不断提升,人工智能在药物筛选领域的应用将更趋智能化与精准化。未来,结合多模态数据的深度学习模型将成为主流,这类模型能够整合分子结构、生物活性、基因组数据乃至临床信息,全面描绘药物作用机制,提升筛选准确率。同时,联邦学习等分布式人工智能技术的发展,有望解决数据隐私和数据共享限制问题,促进跨机构协同创新,推动全球药物研发资源的高效利用,实现更广泛的数据融合和智能决策,助力研发团队在复杂数据环境中做出更准确的判断。

此外,解释性人工智能(XAI)的发展也备受关注。当前许多深度学习模型存在“黑盒”问题,限制了其在药物研发中的应用推广。未来,通过设计更具透明度和可解释性的模型,科研人员能够更好理解算法决策过程,增强模型的可信度和应用安全性。这种可解释性不仅提升了科研人员的信任度,还能帮助发现潜在

的算法偏差和风险,促进算法优化。这将助力监管机构和药企建立更为科学的评价体系,推动人工智能技术在新药研发中的广泛应用,进而加速新药上市进程,满足临床和患者多样化需求,为精准医疗和个性化治疗提供坚实支持,推动整个医药行业迈向更加智能化与高效化的发展阶段。

4.2 面临的主要挑战与对策

尽管人工智能药物筛选技术展现出巨大潜力,但其推广应用仍面临诸多挑战。首先,数据质量和数据量不足仍然是制约算法性能的瓶颈。新药开发过程中,尤其是针对新靶点的药物,缺乏足够的高质量标注数据,导致模型泛化能力受限。解决方案包括构建共享数据库、推动数据标准化及应用迁移学习技术,以缓解数据稀缺问题,促进模型在不同药物类型上的适应性提升。

算法的稳定性和可解释性问题亟待突破。深度模型易受训练数据噪声影响,导致预测结果波动大,缺乏统一标准和验证机制限制了其临床和工业应用。未来应加强算法的鲁棒性研究和解释模型开发,提升 AI 系统在真实环境中的适用性,确保其在复杂药物研发流程中的稳定表现。最后,跨学科人才缺乏与技术整合困难也成为挑战。推动药学、计算机科学与临床医学的深度融合,加强人才培养和技术交

流,打造复合型专业团队,是实现人工智能助力新药研发的关键,只有这样才能真正发挥人工智能的最大效能,推动医药行业持续创新发展。

结论

基于人工智能的药物筛选算法在新药研发中展现出强大的应用潜力,显著提升了药物发现的效率和精准度。通过优化数据预处理、特征工程和模型结构,人工智能技术有效克服了传统筛选方法中的诸多限制,实现了对复杂分子结构的深入理解和高效筛选。虚拟筛选、药效预测及安全性评估等环节的智能化应用,不仅缩短了研发周期,降低了成本,也提升了新药的成功率和安全性。

人工智能药物筛选技术仍面临数据质量不足、模型解释性有限及跨学科协同等多方面挑战。未来的发展需要在多模态数据融合、分布式学习以及解释性人工智能等方面持续创新,以提升算法的鲁棒性和可信度。同时,加强药学与人工智能领域的深度融合,推动标准化数据共享和人才培养,是促进技术落地和产业化的关键。总体来看,人工智能赋能的新药筛选技术正引领药物研发进入智能化时代,有望为解决全球医疗需求和创新药物开发提供坚实支撑。

参考文献

- [1] 张磊. 大数据背景下的 ZD 公司药物研发管理研究[D]. 南京邮电大学, 2022.
- [2] [2] 张斗胜, 许明哲. 计算建模与预测技术在抗生素药物生物学性质研究中的应用进展[J]. 中国新药杂志, 2022, 31(22): 2237-2241.
- [3] 侯春雨. 基于深度学习和神经炎症模型的 de novo 分子设计方法研究[D]. 东北大学, 2021.
- [4] 闵建亮. 基于 2D 分子指纹和非平衡数据集的药物与受体交互作用预测研究[D]. 景德镇陶瓷学院, 2014.
- [5] 张永祥, 周文霞. 中药药理与临床研究进展[M]. 人民卫生出版社: 202302. 946. '