人工智能在高通量药物筛选中的应用研究进展

曹婉昕[1], 张 旭[1], 杨 红[1], 李晓雪[1,2], 李 婉[1], 迟湘胤[1], 赵宇轩[1,3], 杜冠华[1], 王金华[1*]

(1. 中国医学科学院北京协和医学院, 药物研究所, 脑血管病新药发现及晶型研究北京市重点实验室, 北京 100050; 2. 沈阳药科大学, 辽宁 沈阳 117004; 3. 河南大学药学院, 河南 开封 475004)

摘要: 高通量筛选 (high-throughput screening, HTS) 技术在现代药物研发中应用广泛, 能够对化合物进行批量活性测试, 为活性化合物发现提供重要支撑。然而, 其筛选效率和准确性有待提升。为解决这一问题, 人工智能 (artificial intelligence, AI) 技术逐渐融入HTS流程, 通过机器学习和深度学习等算法优势, 优化基于配体的虚拟筛选、联合药物筛选、图像分析等领域的应用, 推动了药物研发的智能化发展。本文回顾了近年来AI技术在HTS中的应用研究进展, 探讨了机器学习模型的应用, 并对AI在HTS相关的化合物筛选和图像识别方面的关键应用进行了总结, 旨在加速创新药物的研发。

关键词: 人工智能; 高通量药物筛选; 深度学习; 虚拟筛选; 图像识别

中图分类号: R966 **文献标识码**: A **文章编号**: 0513-4870(2026)02-0356-09

Advances in the application of artificial intelligence in high-throughput drug screening

CAO Wan-xin[1], ZHANG Xu[1], YANG Hong[1], LI Xiao-xue[1,2], LI Wan[1], CHI Xiang-yin[1], ZHAO Yu-xuan[1,3], DU Guan-hua[1], WANG Jin-hua[1*]

(*1. Beijing Key Laboratory of Innovative Drug Discovery and Polymorphic Druggability Research for Cerebrovascular Diseases, Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100050, China; 2. Shenyang Pharmaceutical University, Shenyang 117004, China; 3. School of Pharmacy, Henan University, Kaifeng 475004, China*)

Abstract: High-throughput screening (HTS) has been widely used in drug discovery recently. It allows batch bioactivity testing of compounds, thereby accelerating the identification of bioactive molecules. However, its screening efficiency and accuracy remain limited. Therefore, artificial intelligence (AI) is gradually integrated into the process of HTS, such as machine learning and deep learning algorithms to enhance ligand-based virtual screening, drug combination screening, and analysis of images, to promote the intelligent development of drug discovery. This review summarized the recent advances in the application of AI technologies in HTS, discussed the applications of machine learning models, and highlighted key AI applications in compound screening and recognition of image in HTS to accelerate the development of innovative drugs.

Key words: artificial intelligence; high-throughput drug screening; deep learning; virtual screening; recognition of image

收稿日期: 2025-06-08; 修回日期: 2025-09-30.

基金项目: 2024年产业技术基础公共服务平台−创新药物筛选验证和检测评价公共服务平台项目 (GN2024-31-4700).

*通讯作者 Tel: 13810230480, E-mail: wjh@imm.ac.cn

DOI: 10.16438/j.0513-4870.2025-0644

高通量筛选 (high-throughput screening, HTS) 技术作为药物发现的重要环节, 是现代药物研发的关键工具, 通过并行化实验设计和自动化操作, 加速了新药候选分子的发现与优化[1-3]。然而, 面对日益庞大的化合物库和复杂的生物靶点网络, 传统HTS技术在数据处理、筛选效率及成本控制方面正面临严峻挑战[4,5]。如何充分挖掘和利用HTS过程中产生的大量多维数据, 已成为提高筛选效率和准确性的重要研究方向。

人工智能 (artificial intelligence, AI) 技术的快速发展为HTS注入了新的活力。通过机器学习 (machine learning, ML) 与深度学习 (deep learning, DL) 的算法优势, AI在基于配体的虚拟筛选、联合药物筛选、高内涵筛选 (high-content screening, HCS) 图像分析及大规模数据处理等方面展现出巨大潜力[6,7]。并且随着图灵奖屡次授予AI领域的先驱, 以及AlphaFold在蛋白质折叠预测中的突破, AI在生物医药领域的应用已变得愈加广泛和重要。这些进展不仅凸显了AI在药物发现中的前瞻性应用, 也推动了HTS技术的跨越式发展。

当前, AI技术与HTS的深度融合不仅显著提升了筛选效率, 也极大地加快了药物研发的可能性。随着算法的不断优化以及计算资源的持续升级, AI在HTS中的应用前景愈加广阔。本文系统回顾了近年来AI技术在HTS领域的应用研究进展, 并结合具体案例深入分析其在不同任务场景中的应用表现。

图1展示了AI技术在HTS中的几种关键应用, 包括化合物筛选、分子生成、高内涵筛选和器官芯片等。

1 化合物筛选

化合物筛选是药物发现流程中的关键环节, 其目的在于从大规模化学库中识别有潜力的活性化合物。AI技术的引入显著提升了筛选效率、准确性及药物发现的速度[8]。

1.1 基于配体的虚拟筛选 基于配体的虚拟筛选 (ligand-based virtual screening, LBVS) 是一种通过计算模拟, 从化合物库中筛选出与已知靶标蛋白质结合能力强的潜在药物分子的方法[9]。这种方法基于已知的配体结构信息, 利用其与靶标的结合模式来预测其他未知化合物与靶标的亲和力。在初步筛选阶段即可帮助筛选出大量苗头化合物用于后续优化。

近年来, AI技术的引入为LBVS提供了更多可能性, 尤其是ML算法和DL算法显著提高了预测的准确性与效率。

传统ML算法在LBVS中应用广泛, 凭借其简单且易于实现的特点, 成为化合物筛选和毒性分析中的常用工具。如支持向量机 (support vector machine, SVM)、随机森林 (random forest, RF) 和K最近邻 (K-nearest neighbors, KNN) 等经典算法, 通过利用化学描述符构建定量构效关系模型, 有效预测了配体与靶标的结合能力[10-12]。这些算法在处理小型和中型数据集时尤其高效, 其分类与回归能力能够迅速识别潜在的活性配体。然而, 传统算法在应对高维、非线性和复杂数据时存在局限, 模型性能容易受到特征选择和数据预处理的影响。此外, 传统算法在面对大规模数据时扩展性不足, 难以充分挖掘多维数据中的潜在模式[13]。

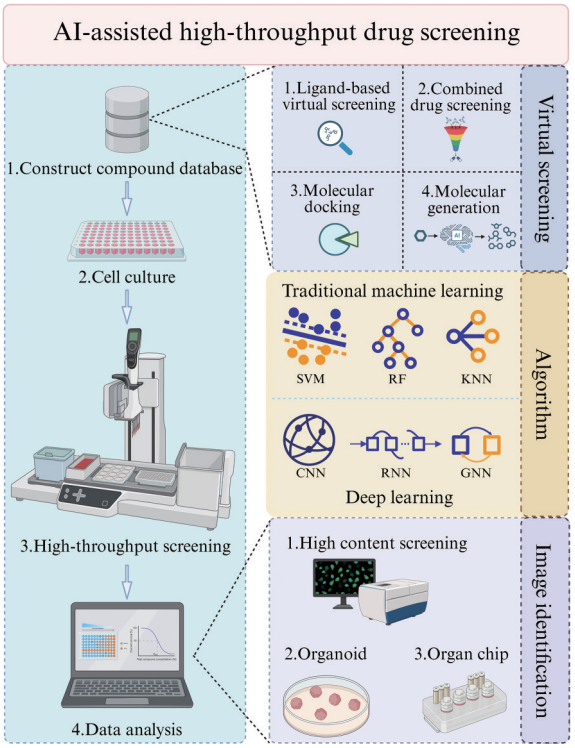


Figure 1 Applications and algorithms of AI-assisted high-throughput drug screening. Created in BioRender. com. SVM: Support vector machine; RF: Random forest; KNN: K-nearest neighbors; CNN: Convolutional neural network; RNN: Recurrent neural network; GNN: Graph neural network

与此不同, DL是ML的一个子领域, 其通过多层级联的非线性处理单元执行特征提取与转换, 构建多层次特征表示并学习抽象概念。相比于传统ML方法, DL在非线性建模和自动特征提取方面表现出更强的优势, 特别是在处理高维和复杂数据时具有显著的能力。DL包含多种神经网络架构, 如卷积神经网络 (convolutional neural network, CNN)、循环神经网络 (recurrent neural network, RNN) 和生成对抗网络 (generative adversarial network, GAN)。这些模型能够自动提取配体和靶标结合过程中的高阶特征, 从而显著提高虚拟筛选的准确性和泛化能力[14]。

近年来, 基于图神经网络 (graph neural network,

GNN) 的 DL 模型在 LBVS 中得到了广泛应用。DeepChem 平台利用 GNN 模型预测化合物的药物代谢和毒性特征。在对比测试中, 该平台在多个关键端点上显著提升了预测性能。DeepChem 模型在代谢稳定性预测任务中 Top-50 准确率超过基线规则系统的最大准确率 (69.5%), 表明其在候选化合物识别中表现优异[15]。通过 AI 模型的辅助, 研究人员能够在早期筛除不符合药代动力学要求的化合物, 从而减少后续筛选的时间和成本。Deep-PK 平台也可利用 GNN 预测小分子药物的药代动力学和毒性。在与药代动力学预测工具 ADMETlab 2.0 的 53 个预测任务比较中, Deep-PK 在 46 个任务上表现优于基线模型, “NR-A”核受体激活端点的马修斯相关系数 (Matthews correlation coefficient, MCC) 提升 0.43, 代谢相关任务的平均 MCC 提升 0.105; 在 36 个毒性预测任务中, Deep-PK 有 30 个端点性能增强, 总体平均 MCC 从 0.58 提升至 0.71[6]。MD-GNN 模型通过将分子结构输入 GNN 并融合实验数据与计算数据, 实现了分子性质预测的性能提升。在 19 种修改后的分子上进行 log*P* 预测时, 模型成功识别出 9 种潜在活性化合物。对比实验显示 MD-GNN 结合特征融合后测试集上的平均绝对误差显著低于仅使用 GNN 或仅使用分子描述符的模型, 分别降低了约 0.2～0.3, 表明特征融合显著提升了预测稳定性和准确性。此外, 与经典 GNN 算法相比, MD-GNN 在测试集上平均误差减少 0.2, 展示了其在多端点预测任务中的优势[16]。FP-GNN 模型结合分子指纹与 GNN, 在 14 条乳腺癌及正常细胞系的表型筛选数据集上平均曲线下面积 (area under the curve, AUC) 达 0.849, 同时在抗噪能力和可解释性方面优于传统分子指纹模型和经典 GNN 模型[17]。Ligandformer 是另一种针对化合物性质预测的先进 AI 模型, 其通过多层自注意力机制从化合物结构中提取局部和全局特征, 并融合不同网络模块的注意力图, 实现了预测结果的稳健性和可解释性。在水溶性、Caco-2 细胞通透性和 Ames 致突变性上, Ligandformer 的预测性能均优于近期设计的 MPNN 和 SAMPN 模型, AUC 分别达到 0.98、0.89 和 0.92。此外, 通过对注意力系数的可视化, 研究者能够直观识别分子中对性质预测贡献最大的原子或片段, 从而为化合物结构优化提供指导[18]。这些模型的成功应用为药物发现提供了新的视角, 加速了药物候选分子的筛选过程。

AI 技术通过学习大量的化合物数据和靶标信息, 自动提取分子结构与生物活性之间的复杂关系。这些方法不仅能够优化配体−靶标相互作用的预测, 还能够在海量数据中发现潜在的活性化合物, 识别出传统方法难以察觉的微弱联系。此外, AI 还能够在分子生成和药物设计中提供指导, 帮助发现新的化合物结构或优化现有药物。AI 与 LBVS 的结合, 不仅加速了药物发现的过程, 也为精准药物设计和个性化治疗提供了强有力的支持。

然而, 这些方法的性能在很大程度上依赖于数据集质量与评估策略。已有研究指出, 不当的数据划分可能导致数据泄漏, 如训练集和测试集存在高度相似的化合物构象或共享受体口袋, 将人为抬高预测准确率[19-22]。此外, 在分子对接与 LBVS 中, 如果未严格控制构象/口袋信息穿越, 模型可能学习到结构先验而非真正的结合规律。交叉验证方式的选择同样会显著影响模型的泛化性能[23-25]。因此, 未来研究需在构建数据集和设计验证流程时更加严谨, 以确保模型性能评价的可靠性与临床可转化性。

1.2 联合药物筛选 联合药物筛选是基于药物组合的策略, 通过组合不同作用机制的药物来增强疗效、克服耐药性或减少不良反应。AI 在联合药物筛选中的应用, 能够帮助识别有效的药物组合。

SynergyGTN 利用 GNN 预测药物组合协同作用, 将药物图特征与细胞系基因表达谱结合。在五折交叉验证中, 受试者工作特征 AUC 和精确率−召回率 AUC 分别达 0.98 和 0.97, 准确率达 0.93, 相比现有最佳方法提升 5%～8%。独立 ASTRAZENECA 数据集验证显示平衡准确率提升 13%, 证明了模型在药物组合筛选中的高准确性和良好泛化能力[26]。GraphSynergy 模型基于蛋白质−蛋白质相互作用网络, 通过图卷积网络和注意机制预测协同抗癌药物组合。该模型在 Drug-CombDB 和 Oncology-Screen 数据集上的准确率分别达到 0.755 3 和 0.755 7, 相比最新发布的 DeepSynergy 算法分别提升了约 12% 和 11%。模型在训练过程中分配高权重的关键蛋白在转录和转录调控等分子功能及生物过程上发挥重要作用, 从而提升了药物组合预测的准确性[27]。

SynergyX 是一个多模态相互作用网络, 通过整合药物和细胞系多组学数据, 实现抗肿瘤药物组合的高精度预测, 比次优模型高出 2%～10%。跨数据集验证显示其稳健性良好, R2 值仅下降约 2%, 并可通过跨模态融合编码器捕捉药物−细胞相互作用, 保证预测可解释性[28]。MatchMaker 模型结合药物化学结构与细胞系基因表达谱, 通过先对单药在特定细胞系中学习嵌入, 再将嵌入向量拼接进行协同效应预测, 实现药物相互作用的量化。相比高通量筛选, 该模型可显著减少实验需求。在对比实验中, MatchMaker 优于 DeepSynergy、TreeCombo、随机森林和 ElasticNet 等传统

ML方法, 体现了其在药物组合协同预测中的优势[29]。

随着这些方法的提出, 研究重点也逐渐从单纯追求预测性能转向模型的临床适用性。一方面, 机制可解释性成为重要考量。许多模型通过注意力权重、图结构池化或特征归因方法, 能够定位到在药物协同中起关键作用的基因或通路, 从而为预测结果提供生物学解释。另一方面, 疗效与毒性的平衡日益受到重视。一些研究在模型中同时引入疗效和毒性相关的表型数据, 或通过多任务学习框架并行预测协同效应和潜在不良反应, 从而避免推荐高效但高毒性的药物组合[30-32]。此外, 多目标优化策略也被引入到联合药物筛选中, 使模型能够在“最大化疗效”与“最小化毒性”之间寻找折中解, 并输出更具临床可行性的候选方案[33-35]。AI不仅显著提升了联合药物筛选的效率和准确性, 也正在通过增强可解释性、引入毒性–疗效权衡及多目标优化等策略, 推动该领域向更具临床转化潜力的方向发展。

1.3 分子对接 分子对接是筛选潜在活性化合物并支持后续先导化合物发现的重要策略之一。近年来, AI驱动的受体–配体对接模型不断发展, 受体–配体对接模型不仅可以通过等变神经网络等方法直接生成复杂的三维原子坐标[36], 还可以学习受体与配体之间距离的概率密度分布, 从而实现结合构象的高效预测[37]。分子对接主要涉及基于蛋白质三维结构、分析靶蛋白结合位点特征及其与小分子药物的相互作用, 采用亲和力评分函数评估蛋白与药物的结合亲和力, 从大量的化合物分子中选择具有高预测评分的候选化合物用于后续的生物活性测试。如DL对接平台DeepDocking结合FRED对接程序时, 可以快速准确地计算出ZINC15库中13.6亿个分子与12个主要靶蛋白的对接分数, 并证明了高达100倍的数据缩减和6 000倍的高评分分子的富集[38]。DL算法KDBNet结合了3D蛋白质和分子结构数据来预测结合亲和力, 实验表明其在预测激酶药物亲和力方面优于现有的DL模型。当与贝叶斯优化框架集成时, KDBNet实现了数据高效的主动学习, 并加速了对各种高结合激酶药物对的探索和利用[39]。也有研究报道主动学习和贝叶斯优化等技术是解决化学空间搜索问题的有效方法, 成为提高筛选效率的关键[40]。

早期ML在亲和力预测方面的成功激发了人们对DL模型的兴趣, 而且这些模型处理三维结构和非结构数据的性能可能优于传统的评分函数。如一种通用三维分子表征学习框架Uni-Mol显著扩展了分子表征能力和应用范围。在蛋白质–配体结合构象预测任务中, 其预测均方根偏差可低至1.23 Å, 而在分子构象生成任务中, 成功率超过90%[41]。除了利用GNN外, DiffDock将DL与图卷积网络相结合, 实现蛋白–配体盲对接预测, 显著提高了分子对接的准确性和计算速度。该方法在评估全氟烷基和多氟烷基物质与血液蛋白结合亲和力方面显示出高效性能, 为研究其在人体内的命运及潜在分子作用路径提供了快速手段[42]。RTMScore是一种新的评分函数, 其通过混合密度网络建立残基–原子距离似然势模型。在CASF基准的对接测试中, RTMScore在Top1对接成功率上达到93.1%和89.7%, 显著优于大多数DL方法和传统方法。此外, RTMScore在正向筛选任务中, 可在排名前1%、5%和10%的候选分子中, 分别识别出66.7%、83.6%和90.1%的最高亲和力配体; 并且在反向筛选任务中也保持了明显优势。RTMScore在对接准确性和筛选能力上均展现出优于现有主流方法的综合性能[41]。PLANET模型结合蛋白结合口袋的三维结构和配体二维结构。在主流公开数据集的评估中, PLANET达到了与传统对接程序Glide相当的精度, 但它花费的计算时间不到Glide计算时长的1%[43]。这些AI参与的分子对接算法改进有效地提高了其性能, 具有良好的应用前景。

1.4 分子生成与大模型驱动的设计 随着生成式AI的快速发展, 分子生成与优化逐渐成为药物筛选的重要补充手段。不同于传统的基于配体或结构的虚拟筛选, 这类方法能够直接在化学空间中生成具有特定性质的新分子, 从而实现“设计–筛选”一体化, 加速活性化合物的发现。

扩散模型的兴起显著推动了分子生成领域的研究。与变分自编码器、生成对抗网络等早期方法相比, 扩散模型在建模复杂分子分布时表现出更强的稳定性与表达能力。如SE(3)-等变扩散模型能够保持三维分子结构在旋转与平移下的不变性, 从而在蛋白–配体复合物的三维生成和柔性对接任务中取得了优异表现[44]。此外, 跨尺度蛋白–配体建模还可直接生成与靶蛋白结合口袋匹配的小分子结构, 通过引入多目标优化策略, 同时考虑分子的合成可行性和对接分数, 模型的可合成性分数可提升至0.78, 相比传统方法和未优化模型具有明显优势[45]。

大规模语言模型也被引入分子序列与图结构的建模中。3D-MolT5通过统一编码分子的一维序列、三维结构和文本信息, 实现了跨模态理解。在下游任务中, 包括分子性质预测、三维分子描述生成和文本驱动分子生成, 3D-MolT5均优于传统基线模型[46]。ChemLLM则进一步将多模态能力扩展至分子图像处理, 可同时执行分子图像描述生成、图像性质预测、图

像到SMILES转换以及多目标分子图像设计等任务，在这些任务中表现优于现有多模态大模型，并兼具较强的任务通用性[47]。此外，近期提出的多模态化学与生物大模型能够同时处理分子结构与蛋白质序列，甚至整合分子图像等信息，实现跨尺度的相互作用预测[48,49]。此类方法不仅提升了分子筛选与结合预测的准确性，还在复杂任务如抗体−抗原相互作用和蛋白−小分子协同设计中展现出潜力。

2 图像识别

图像识别技术与DL算法结合能够高效处理大量实验数据，推动HTS的自动化，目前已广泛应用于HCS、类器官和器官芯片技术。

2.1 高内涵筛选 自1997年Cellomics公司推出首个用于HTS的高内涵成像平台以来，HCS技术已成为药物研发领域的核心技术之一。HCS是一种基于细胞图像分析的HTS技术，通过结合DL与多组学数据分析方法，显著提升了细胞表型分类的效率和预测的准确性[50]。传统的HCS方法多采用二维细胞培养和专有算法进行图像分析，但随着开源软件CellProfiler、Definiens、Advanced Cell Classifier和DL技术的发展，图像分析从简单的定量测量扩展到复杂细胞表型的深度分析。通过图像分割和特征提取算法，这些软件能够从复杂的图像中提取细胞的形态学特征，为药物发现提供了强有力的支持[51]。

通过将HCS与基因组学、蛋白质组学等多组学数据结合，研究人员能够更全面地了解细胞对药物的反应和作用机制，从而加速新疗法的发现。利用Image2Omics DL方法，研究人员成功从高内涵图像中预测了M1和M2型巨噬细胞中部分转录本和蛋白质的丰度，证明细胞成像可能可作为多组学测量的可扩展替代方法[50]。此外，基于YOLOv5的DL检测在巨噬细胞HCS中表现出优异的性能，在高细胞密度区域显著提高了检测精度，运行速度提升了约2倍[52]。跨域一致性学习方法被提出用于解决批次效应问题，在没有高质量标签的情况下，通过消除批次特定的噪声，提升了细胞表型的分类精度，进而提高了治疗效果的辨识能力[53]。HCS为表型组学提供了重要支持，帮助揭示细胞在治疗中的潜在变化，为药物筛选提供了新的方向。

有研究人员基于LPS刺激的小鼠巨噬细胞构建DL形态学筛选平台，通过7种荧光探针标记细胞表型，训练深度神经网络区分正常与炎症状态，从2 259种已知机制的化合物中筛选出12个能显著调控LPS诱导细胞因子表达的活性化合物，其中7种为MEK1/2特异性抑制剂。该高内涵分析方法为抗炎药物发现提供了新策略[54]。此外，DL平台FociNet可自动分割全场荧光图像并剖析每个细胞的DNA损伤。经6 000张单核图像注释训练后，FociNet能较好地完成对细胞的正常、受损或无信号状态分类。它还用于分析315种中药天然化合物HCS的5 000多张病灶图像数据集，成功鉴定出依伏地胺等几种能显著减少DNA损伤标志的新型活性化合物，其中异甘草素能显著减少双链断裂。该研究提供了基于显微镜评估DNA损伤的AI辅助新药发现的新思路[55]。

在HCS中，数据质量对模型性能至关重要。典型端到端管线包括原始图像采集、标注策略、预处理和特征提取等环节。标注策略可采用手工注释、半监督学习或基于已知表型的自动标注，以保证训练数据的准确性[56]。批次效应是高通量实验中常见问题，不同实验批次、成像条件或染色批次可能引入系统性偏差，可通过批次标准化、对抗性训练或跨域一致性学习等方法进行矫正[57,58]。

2.2 类器官 类器官是一种来源于干细胞或组织特异性祖细胞的三维培养系统，能够在体外模拟器官的复杂结构和功能。因此，它们被广泛用于疾病建模、个性化医疗和药物筛选。然而，传统的类器官分析方法通常依赖于手动评估或低效的计算方法，难以满足HTS的需求。DL技术在图像分割、目标检测和时空分析等领域的快速发展，使其在类器官研究中展现出巨大潜力。基于DL技术的AI为类器官的自动化检测、形态学分析、功能评估和药物筛选提供了强大的支持。

类器官的大小、形状、结构完整性等形态特征是评估其健康状态、增殖能力和药物反应的重要指标。然而，类器官的三维培养环境使得成像过程中容易受到焦平面、光散射和样本异质性的影响，从而增加了形态分析的难度。近年来，DL技术在类器官形态分析中展现出卓越的性能。如OrganelX和OrgaSegment采用基于DL的分割方法，实现了对小鼠肝类器官和囊性纤维化类器官的自动检测和形态学量化，提高了测量的准确性和一致性[59]。此外，Deep-Orga利用轻量级神经网络RepVGG优化了计算效率，在中型和小型类器官检测中分别实现了74%和48%以上的平均精度均值，为高通量分析提供了可行方案[60]。

除了形态特征，类器官的生长速率、分化趋势和细胞死亡等动态行为也是衡量其生物学功能的重要参数。传统的ATP测定方法只能提供整体生长活力信息，无法解析单个类器官的响应。基于AI的视频分析方法，如使用SAM和DINOv2的时空分析模型，能够从长时间序列显微镜数据中提取高维特征，实现对类器官药物响应的精准预测，在4个验证集上的平均预

测误差仅为0.175 5[61]。此外, Cellos将传统计算方法与DL结合, 能够在三维类器官中区分不同类型的细胞, 并定量分析其空间分布。在对混合荧光标记的类器官中, Cellos可准确计算不同克隆细胞的比例, 并追踪药物处理下的克隆特异性响应, 进一步提升了类器官功能评估的精细化程度[62]。

随着类器官模型在HTS中的广泛应用, 如何高效解析其形态与响应特征成为关键问题。基于AI的HTS策略不仅能提高实验效率, 还能优化筛选流程。如3D-Organoid-SwinNet结合Transformer模型, 实现了更精确的类器官分割, 验证集上的Dice指数高达94.9%, 且模型仅含约2 100万个参数, 为蛋白质读取和药物筛选提供了更高效的解析工具[63]。此外, ACU2Net通过改进U-Net结构, 在膀胱癌类器官药物筛选中显著提升分割精度, 其敏感性、特异性和F1值分别达94.81%、88.50%和91.54%, 为抗癌药物效果评估提供了可靠工具[64]。

2.3 器官芯片 器官芯片是基于微流体细胞培养的仿生体外系统, 与传统的二维细胞培养或动物模型相比, 器官芯片具有更高的生理相关性, 并能提供动态的实验环境, 更加真实地模拟人体器官的生理状态和疾病过程。器官芯片的HTS往往依赖端到端的数据处理管线, 包括图像采集、预处理、分割与特征提取、动态行为分析以及预测模型训练等环节, 从而确保能够将原始成像数据系统化地转化为功能性结论。

如何维持稳定且动态的细胞培养环境是器官芯片领域的一大难题, 尤其是在多器官系统 (multiple organ system, MOS) 模型中, 其环境的稳定性和动态调控更具挑战性。由于人体是一个复杂的生理系统, MOS通过微型通道集成多个细胞培养室, 以模拟药物作用下的整体生理功能, 从而更全面地解析药物的剂量反应和毒性机制[65]。因此, 确保器官芯片在动态培养条件下的稳定性对于模拟真实的体内环境至关重要。DL等AI技术已广泛应用于分析来自器官芯片传感器的多维数据, 这些传感器可实时监测pH值、氧气水平、代谢活动乃至细胞形态等关键生理参数, 以优化实验条件, 提升数据的准确性和实验的可重复性[66]。

智能器官芯片系统 (intelligent organ-on-a-chip, iOOC) 能够实时反馈动态调节电、化学或机械刺激, 以维持器官模型的稳态, 从而更准确地模拟生理系统中的反馈调节机制。通过借鉴工程学中的闭环控制策略, iOOC可以集成生物传感器监测关键生理参数, 将测得的数据与目标状态进行比较, 以计算误差并调整刺激输入。常见的控制方法包括开/关控制、比例–积分–微分控制, 以及结合ML的预测性优化策略[67]。AI与器官芯片之间的实时反馈循环不仅加速了潜在活性化合物的发现, 还帮助研究人员在器官层面深入理解疾病机制, 推动了临床前研究向临床应用的转化。

CNN与光学显微技术及其他成像技术的结合, 显著提高了对器官芯片中细胞行为与形态变化的分析效率[68]。AI驱动的图像识别系统能够自动化地检测药物处理后的细微变化, 如细胞形态变化或膜完整性损伤, 而这些变化往往是药物毒性或疗效的重要指标, 从而为HCS提供了更加精确的信息[69]。

3 总结与展望

AI技术在HTS中的应用, 极大地推动了药物研发的智能化进程。随着ML和DL技术的不断发展, AI在化合物筛选、图像分析和数据处理等方面优势显著。AI技术不仅优化了传统HTS方法的局限性, 还在联合药物筛选和多模态数据融合中提供了新的解决方案。如基于GNN的化合物性质预测模型, 成功提升了化合物的药物代谢、毒性预测等关键指标的准确性, 为潜在活性化合物和先导化合物的发现与优化提供了重要参考。DL方法还被广泛应用于药物–靶标相互作用预测、分子生成以及HCS图像分析中, 进一步加速了药物发现的步伐。

尽管AI技术在HTS领域应用广泛, 但仍存在一些局限性。在药物发现和生物医学领域, AI模型需要处理高维复杂数据, 导致计算资源消耗巨大[70]。虽然DL通过自动特征提取和非线性建模显著提高了预测精度, 但其模型训练过程不仅需要大量标注数据, 还伴随着高昂的计算成本[71]。此外, AI模型的可解释性问题和透明度也严重制约了其在药物筛选和生物活性预测领域的进展[72]。

随着计算能力的持续提升和分布式计算技术的发展, 未来AI模型的训练成本有望降低。云计算和边缘计算的深度融合将为大规模数据处理提供更强大的技术支持[73]。在模型发展方面, 可解释性AI将成为重点研究方向, 开发透明、可追溯的算法将有助于增强AI在药物研发中的可信度和实用性。在药物组合筛选领域, AI将在应对抗药性问题和复杂疾病治疗方面发挥更大作用, 为创新药物组合方案的设计提供有力支持。此外, 多模态学习和跨领域数据融合技术的突破将推动HTS向更高层次发展, 使AI技术在海量数据的深度挖掘与综合分析中展现出更大的应用价值。随着技术不断发展和优化, AI将在药物发现和疾病治疗领域得到更广泛的应用, 进一步推动个性化治疗和精准医疗的发展。

作者贡献: 曹婉昕、张旭负责论文撰写与图片绘制; 杨

红、李晓雪、李婉、迟湘胤、赵宇轩负责文献收集与分析; 王金华负责论文的选题、构思并指导文章的撰写; 王金华、杜冠华负责稿件的修改与校对。所有作者阅读并认可终稿。

利益冲突: 本文作者声明无利益冲突。

References

[1] Jiang J, Li H, Xuan Y, et al. High-throughput discovery of novel anti-colorectal cancer active compounds targeting TRAF2 [J]. Acta Pharm Sin (药学学报), 2025, 60: 1721-1729.

[2] You BQ, Zhou WW, Li Y, et al. Establishment and application of a cell-based high-throughput screening model for TMPRSS2 inhibitors [J]. Acta Pharm Sin (药学学报), 2024, 59: 3273-3281.

[3] Yan GG, Yan HH, Liu ZC, et al. Optimization and application of a colorimetric high-throughput screening model for small-molecule inhibitors of SARS-CoV-2 main protease [J]. Acta Pharm Sin (药学学报), 2022, 57: 3019-3026.

[4] Blucher AS, McWeeney SK. Challenges in secondary analysis of high throughput screening data [J]. Pac Symp Biocomput, 2014. DOI: 10.1142/9789814583220_0012.

[5] Roy A. Challenges with risk mitigation in academic drug discovery: finding the best solution [J]. Expert Opin Drug Discov, 2019, 14: 95-100.

[6] Myung Y, De Sá AGC, Ascher DB. Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction [J]. Nucleic Acids Res, 2024, 52: W469-W475.

[7] Carreras-Puigvert J, Spjuth O. Artificial intelligence for high content imaging in drug discovery [J]. Curr Opin Struct Biol, 2024, 87: 102842.

[8] Li ZY, Cong KY, Wu SQ, et al. Advances in the application of artificial intelligence technology in lead compound discovery and optimization [J]. Acta Pharm Sin (药学学报), 2024, 59: 2443-2453.

[9] Liu RZ, Song JK, Liu AL, et al. Advances in the application of artificial intelligence in ligand- and receptor-based drug screening [J]. Acta Pharm Sin (药学学报), 2021, 56: 2136-2145.

[10] Rudresh BB, Tater AK, Barot V, et al. Development and experimental validation of 3D QSAR models for the screening of thyroid peroxidase inhibitors using integrated methods of computational chemistry [J]. Heliyon, 2024, 10: e29756.

[11] Li R, Tian Y, Yang Z, et al. Classification models and SAR analysis on HDAC1 inhibitors using machine learning methods [J]. Mol Divers, 2023, 27: 1037-1051.

[12] Orsi M, Shing Loh B, Weng C, et al. Using machine learning to predict the antibacterial activity of ruthenium complexes [J]. Angew Chem Int Ed Engl, 2024, 63: e202317901.

[13] Dara S, Dhamercherla S, Jadav SS, et al. Machine learning in drug discovery: a review [J]. Artif Intell Rev, 2022, 55: 1947-1999.

[14] Du X, Chen Z, Li Q, et al. Organoids revealed: morphological analysis of the profound next generation *in-vitro* model with artificial intelligence [J]. Biodes Manuf, 2023, 6: 319-339.

[15] Ramsundar B. Molecular Machine Learning with Deepchem [D]. San Francisco: Stanford University, 2018.

[16] Chen S, Wulamu A, Zou Q, et al. MD-GNN: a mechanism-data-driven graph neural network for molecular properties prediction and new material discovery [J]. J Mol Graph Model, 2023, 123: 108506.

[17] Cai H, Zhang H, Zhao D, et al. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction [J]. Brief Bioinform, 2022, 23: bbac408.

[18] Guo J, Liu Q, Guo H, et al. Ligandformer: a graph neural network for predicting compound property with robust interpretation [J]. arXiv, 2022. DOI:10.48550/arXiv.2202.10873.

[19] Papernot N, McDaniel P, Goodfellow I. Transferability in machine learning: from phenomena to black-box attacks using adversarial samples [J]. arXiv, 2016. DOI: 10.48550/arXiv.1605.07277.

[20] Schwaller P, Laino T, Gaudin T, et al. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction [J]. ACS Cent Sci, 2019, 5: 1572-1583.

[21] Wang S, Sun S, Li Z, et al. Accurate *de novo* prediction of protein contact map by ultra-deep learning model [J]. PLoS Comput Biol, 2017, 13: e1005324.

[22] Tran-Nguyen VK, Jacquemard C, Rognan D. LIT-PCBA: an unbiased data set for machine learning and virtual screening [J]. J Chem Inf Model, 2020, 60: 4263-4273.

[23] Guo Q, Hernandez-Hernandez S, Ballester PJ. Scaffold splits overestimate virtual screening performance [C] // Proceedings of The International Conference on Artificial Neural Networks. Cham: Springer, 2024: 58-72.

[24] Li J, Guan X, Zhang O, et al. Leak proof PDBBind: a reorganized dataset of protein-ligand complexes for more generalizable binding affinity prediction [J]. arXiv, 2024. DOI: 10.48550/arXiv.2308.09639.

[25] Moshawih S, Bu ZH, Goh HP, et al. Consensus holistic virtual screening for drug discovery: a novel machine learning model approach [J]. J Cheminform, 2024, 16: 62.

[26] Alam W, Tayara H, Chong KT. Unlocking the therapeutic potential of drug combinations through synergy prediction using graph transformer networks [J]. Comput Biol Med, 2024, 170: 108007.

[27] Yang J, Xu Z, Wu WKK, et al. Graphsynergy: a network-inspired deep learning model for anticancer drug combination prediction [J]. J Am Med Inform Assoc, 2021, 28: 2336-2345.

[28] Guo Y, Hu H, Chen W, et al. Synergyx: a multi-modality mutual attention network for interpretable drug synergy prediction [J]. Brief Bioinform, 2024, 25: bbae015.

[29] Kuru HI, Tastan O, Cicek AE. Matchmaker: a deep learning framework for drug synergy prediction [J]. IEEE/ACM Trans

Comput Biol Bioinform, 2022, 19: 2334-2344.

[30] Monem S, Hassanien AE, Abdel-Hamid AH. A multi-task graph deep learning model to predict drugs combination of synergy and sensitivity scores [J]. BMC Bioinformatics, 2024, 25: 327.

[31] Shan W, Shen C, Luo L, et al. Multi-task learning for predicting synergistic drug combinations based on auto-encoding multi-relational graphs [J]. iScience, 2023, 26: 108020.

[32] Chen D, Wang X, Zhu H, et al. Predicting anticancer synergistic drug combinations based on multi-task learning [J]. BMC Bioinformatics, 2023, 24: 448.

[33] Gevertz JL, Kareva I. Guiding model-driven combination dose selection using multi-objective synergy optimization [J]. CPT Pharmacometrics Syst Pharmacol, 2023, 12: 1698-1713.

[34] Arora HS, Lev K, Robida A, et al. A mechanistic neural network model predicts both potency and toxicity of antimicrobial combination therapies [J]. medRxiv, 2025. DOI: 10.1101/2025.03.19.25324270.

[35] Nicolotti O, Giangreco I, Introcaso A, et al. Strategies of multi-objective optimization in drug discovery and development [J]. Expert Opin Drug Discov, 2011, 6: 871-884.

[36] Qiao ZR, Nie WL, Vahdat A, et al. State-specific protein-ligand complex structure prediction with a multiscale deep generative model [J]. Nat Mach Intell, 2024, 6: 195-208.

[37] Méndez-Lucio O, Ahmad M, Del Rio-Chanona EA, et al. A geometric deep learning approach to predict binding conformations of bioactive molecules [J]. Nat Mach Intell, 2021, 3: 1033-1039.

[38] Gentile F, Agrawal V, Hsing M, et al. Deep docking: a deep learning platform for augmentation of structure based drug discovery [J]. Acs Central Sci, 2020, 6: 939-949.

[39] Luo YA, Liu Y, Peng J. Calibrated geometric deep learning improves kinase-drug binding predictions [J]. Nat Mach Intell, 2023, 5: 1390-1401.

[40] Cao ZL, Sciabola S, Wang Y. Large-scale pretraining improves sample efficiency of active learning-based virtual screening [J]. J Chem Inf Model, 2024, 64: 1882-1891.

[41] Zhou G, Gao Z, Ding Q, et al. Uni-Mol: a universal 3D molecular representation learning framework [C] // Proceedings of The Eleventh International Conference on Learning Representations. Kigali Rwanda: International Conference on Learning Representations, 2023.

[42] B Fortela DL, Mikolajczyk AP, Carnes MR, et al. Predicting molecular docking of per- and polyfluoroalkyl substances to blood protein using generative artificial intelligence algorithm diffdock [J]. Biotechniques, 2024, 76: 14-26.

[43] Zhang XY, Gao HT, Wang HJ, et al. Planet: a multi-objective graph neural network model for protein-ligand binding affinity prediction [J]. J Chem Inf Model, 2023, 64: 2205-2220.

[44] Zhu J, Gu Z, Pei J, et al. DiffBindFR: an SE (3) equivariant network for flexible protein-ligand docking [J]. Chem Sci, 2024, 15: 7926-7942.

[45] Cremer J, Le T, Noé F, et al. Pilot: equivariant diffusion for pocket-conditioned *de novo* ligand generation with multi-objective guidance *via* importance sampling [J]. Chem Sci, 2024, 15: 14954-14967.

[46] Pei Q, Wu L, Gao K, et al. 3D-MolT5: towards unified 3D molecule-text modeling with 3D molecular tokenization [J]. arXiv, 2024. DOI: 10.48550/arXiv.2406.05797.

[47] Zhang D, Liu W, Tan Q, et al. ChemLLM: a chemical large language model [J]. arXiv, 2024. DOI: 10.48550/arXiv.2402.06852.

[48] Luo Y, Liu XY, Yang K, et al. Toward unified AI drug discovery with multimodal knowledge [J]. Health Data Sci, 2024, 4: 0113.

[49] Karampuri A, Kundur S, Perugu S. Exploratory drug discovery in breast cancer patients: a multimodal deep learning approach to identify novel drug candidates targeting RTK signaling [J]. Comput Biol Med, 2024, 174: 108433.

[50] Mehrizi R, Mehrjou A, Alegro M, et al. Multi-omics prediction from high-content cellular imaging with deep learning [J]. arXiv, 2023. DOI: 10.48550/arXiv.2306.09391.

[51] Way GP, Sailem H, Shave S, et al. Evolution and impact of high content imaging [J]. SLAS Discov, 2023, 28: 292-305.

[52] Ong K, Cai X, Marur V, et al. Deep learning-based rapid macrophage cell detection and localization in high-content microscopy screening [C] // Proceedings of The Medical Imaging 2023: Digital and Computational Pathology. Bellingham: SPIE, 2023: 310-316.

[53] Haslum JF, Matsoukas C, Leuchowius KJ, et al. Metadata-guided consistency learning for high content images [C]// Proceedings of the Medical Imaging with Deep Learning 2024. Hong Kong: Proceedings of Machine Learning Research, 2024: 918-936.

[54] Lau TA, Mair E, Rabbitts BM, et al. High-content image-based screening and deep learning for the detection of anti-inflammatory drug leads [J]. Chembiochem, 2024, 25: e202300136.

[55] Chen X, Xun D, Zheng R, et al. Deep-learning-assisted assessment of DNA damage based on foci images and its application in high-content screening of lead compounds [J]. Anal Chem, 2020, 92: 14267-14277.

[56] Teikari P, Santos M, Poon C, et al. Deep learning convolutional networks for multiphoton microscopy vasculature segmentation [J]. arXiv, 2016. DOI: 10.48550/arXiv.1606.02382.

[57] Arevalo J, Su E, Ewald JD, et al. Evaluating batch correction methods for image-based cell profiling [J]. Nat Commun, 2024, 15: 6516.

[58] Marée R. The need for careful data collection for pattern recognition in digital pathology [J]. J Pathol Inform, 2017, 8: 19.

[59] Lefferts JW, Kroes S, Smith MB, et al. OrgaSegment: deep-learning based organoid segmentation to quantify CFTR dependent fluid secretion [J]. Commun Biol, 2024, 7: 319.

[60] Leng B, Jiang H, Wang B, et al. Deep-Orga: an improved deep learning-based lightweight model for intestinal organoid detection [J]. Comput Biol Med, 2024, 169: 107847.

[61] Fillioux L, Gontran E, Cartry J, et al. Spatio-temporal analysis of patient-derived organoid videos using deep learning for the prediction of drug efficacy [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: Institute of Electrical and Electronics Engineers / Computer Vision Foundation, 2023: 3930-3939.

[62] Mukashyaka P, Kumar P, Mellert DJ, et al. Abstract A032: Cellos: high-throughput deconvolution of 3D organoid dynamics at cellular resolution for cancer pharmacology [J]. Cancer Res, 2024, 84: A032.

[63] Sohaib M, Shabani S, Mohammed SA, et al. 3D-organoid-swinnet: high-content profiling of 3D organoids [J]. IEEE J Biomed Health Inform, 2025, 29: 792-798.

[64] Zhang S, Li L, Yu P, et al. A deep learning model for drug screening and evaluation in bladder cancer organoids [J]. Front Oncol, 2023, 13: 1064548.

[65] Joseph X, Akhil V, Arathi A, et al. Comprehensive development in organ-on-a-chip technology [J]. J Pharm Sci, 2022, 111: 18-31.

[66] Deng S, Li C, Cao J, et al. Organ-on-a-chip meets artificial intelligence in drug evaluation [J]. Theranostics, 2023, 13: 4526-4558.

[67] Li X, Zhu H, Gu B, et al. Advancing intelligent organ-on-a-chip systems with comprehensive *in situ* bioanalysis [J]. Adv Mater, 2024, 36: e2305268.

[68] Dai M, Xiao G, Shao M, et al. The synergy between deep learning and organs-on-chips for high-throughput drug screening: a review [J]. Biosensors (Basel), 2023, 13: 389.

[69] Sabaté Del Río J, Ro J, Yoon H, et al. Integrated technologies for continuous monitoring of organs-on-chips: current challenges and potential solutions [J]. Biosens Bioelectron, 2023, 224: 115057.

[70] Zhang Y, Liu C, Liu M, et al. Attention is all you need: utilizing attention in AI-enabled drug discovery [J]. Brief Bioinform, 2023, 25: bbad467.

[71] Li J, Chen J, Bai H, et al. An overview of organs-on-chips based on deep learning [J]. Research (Wash D C), 2022, 2022: 9869518.

[72] 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.

[73] Atieh AT. The next generation cloud technologies: a review on distributed cloud, fog and edge computing and their opportunities and challenges [J]. ResearchBerg Rev Sci Technol, 2021, 1: 1-15.