

深度学习在药物研发中的研究进展

廖俊¹, 徐洁洁¹, 皮志鹏¹, 窦智扬¹, 尚靖^{2*}

(1. 中国药科大学理学院, 江苏 南京 211198; 2. 中国药科大学中药学院, 江苏 南京 211198)

[摘要] 近年来以深度学习为代表的人工智能技术与医学、药学等多个领域深度融合。深度学习被应用于蛋白质结构与功能预测、药物靶点预测、药物代谢动力学性质预测、药物有效性及安全性预测以及药物相互作用预测等多个药物研发环节, 取得了显著成就, 提高研发效率的同时降低临床前试验以及临床试验相关的成本和风险。通过总结多种深度学习方法在药物研发各个过程中的具体应用及分析不同深度学习方法在药物研发中的应用特点, 阐述了深度学习在药物研发中现存的一些问题并做出展望, 以期为进一步研究提供借鉴的思路和方法。

[关键词] 深度学习; 药物研发; 新机制和新靶点; 药物相互作用; 药物不良反应

[中图分类号] R9-39

[文献标志码] A

[文章编号] 1001-5094 (2020) 05-0387-08

Advances in the Application of Deep Learning in Drug Development

LIAO Jun¹, XU Jiejie¹, PI Zhipeng¹, DOU Zhiyang¹, SHANG Jing²

(1. School of Science, China Pharmaceutical University, Nanjing 211198, China; 2. School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, China)

[Abstract] Recently, artificial intelligence (AI) represented by deep learning (DL) has been deeply integrated with various fields of medical and pharmaceutical sciences. DL has been applied to drug research and development with significant progress in the prediction of protein structure and function, drug target, pharmacokinetic properties, drug safety and efficacy, and drug interactions, with improved efficiency of drug research and development and reduced costs and risks of preclinical and clinical trials. This review summarizes the specific applications of various methods of DL in the whole process of drug research and development, with an analysis of the application characteristics of different DL methods in drug research and development. Finally, some problems and prospects of DL in drug research and development are presented so as to provide reference for ideas and methods in further research.

[Key words] deep learning; drug research and development; new mechanism and new target; drug-drug interaction; adverse drug reaction

药物研发过程主要包括药物靶点确定、先导化合物的发现与优化、候选药物确定、临床前研究和临床研究^[1]。整个药物研发进程, 就是在验证某个靶点在人体中的生物学功能的过程。而药物靶点的缺乏、动物模型临床转化差、疾病异质性及生物系统内在的复杂性等问题, 使得药物研发成为一个漫长而又艰难的过程。开发一种新型处方药, 平均税前支出约为 25.58 亿美元^[2], 大约需要 10~15 年。尽管投入成本高, 但在药物研发过程中创新小分子的临床批准成功率却只有 13%, 失败的风险相对较高。早期药物发现主要基于经验尝试, 化合物筛选以及偶然发现获得。在现代药物研发中, 机器学习

在定量结构-活性关系 (quantitative structure activity relationship, QSAR) 模型^[3]、定量结构-性质关系 (quantitative structure property relationship, QSPR) 模型^[4]等方面发挥重要作用。不同于传统的机器学习方法使用手动设计的特征, 最新的深度学习 (deep learning, DL) 方法可以自动从输入数据中学习特征, 通过多层特征提取将低层特征转换为高层特征。由于其强大的泛化和特征提取能力, 将其应用在药物开发的不同环节, 包括蛋白质结构与功能预测、药物代谢动力学性质预测、药物有效性及安全性预测以及药物相互作用预测等, 显示出巨大的前景。本综述回顾了近几年来 DL 在药物研发中的应用, 并对当前问题提出建议以及展望。

接受日期: 2020-01-01

项目资助: 国家自然科学基金资助项目 (No.81874331); 双一流创新团队资助项目 (No. CPU2018GY19)

***通讯作者:** 尚靖, 教授;

研究方向: 中药药理;

Tel: 025-51190631; **E-mail:** shangjing21cn@163.com

1 深度学习与药物研发数据

DL 的概念由 Hinton 等^[5]于 2006 年提出, 其概念源于人工神经网络的研究。DL 的结构是一种含多

隐层的多层感知器结构, 其通过组合底层特征形成更加抽象的高层来表示属性类别或特征, 以发现数据的分布式特征表示。DL 理论中包含了许多不同的深度神经网络模型, 例如经典的深层神经网络 (deep neural network, DNN)、深层置信网络、卷积神经网络 (convolutional neural network, CNN)、深层玻尔兹曼机 (deep boltzmann machines, DBM)、循环神经网络 (recurrent neural network, RNN) 等。不同结构的网络适用于处理不同的数据类型, 例如 CNN 适用于图像处理, RNN 适用于语音识别等。同时, 通过与不同算法的联用这些网络模型还会产生一些不同的变种。

目前正处在医药产业发展的关键节点, 由于新药物靶点和作用机制的发现越来越难, 新药研发需要投入更多的资金和精力。提升研发效率和深度挖掘已有数据来发现新的规律是解决该问题的有效途径之一, 而 DL 在这 2 个方面都可以有广泛的应用, 因此许多制药公司和药物研发机构都将 DL 方法用于辅助药物研发。例如: Berg 公司基于人工智能的 Interrogative Biology 平台技术^[6]通过分析海量病人和正常人样本 (如蛋白相互作用网络) 来寻找治疗疾病的新靶点和诊断疾病的生物标志物, 以 Berg 公司进行肿瘤药物研究为例, 通过收集大量生物样本, 如血液、肿瘤组织或肿瘤患者的尿液, 同时也收集捐助者的健康组织样本。研究人员会创建细胞株, 然后将其放进不同的模拟患者发病时的实际状态环境下进行观察, 有比如低氧环境, 高血糖患者细胞及肿瘤细胞喜欢生活的环境。细胞株建立之后, 对其中的基因、蛋白质、代谢物和脂肪进行标识并形成节点, 不同节点的重要程度不同, 重要程度越高与疾病的关联越大。Narain 等^[7]曾基于此平台介导发现胰腺癌的检测、分层和预后的分子标记; IBM Watson 为 IBM 旗下的认知计算系统, 技术平台。认知计算代表一种全新的计算模式, 它包含信息分析, 自然语言处理和机器学习领域的大量技术创新。该新药发现系统^[8]通过自然语言处理技术分析海量文献, 寻找潜在的关联性来预测新的假说推动新药研发; Engine Biosciences^[9], 也是利用人工智能技术来进行老药新用、新靶点开发以及精准医疗等服务

的互联网产品。图 1 列举了目前 DL 在药物研发不同阶段已经取得成果的相关应用。

2 深度学习在药物研发中的应用

2.1 蛋白质结构与功能

蛋白质的功能研究在生命科学中占据重要的地位, 大多数疾病的发生都与蛋白质功能障碍有关。1973 年, Anfinsen^[10]发现变性的只保留了一级结构的核糖核酸酶可以重新折叠并恢复生物活性, 说明代表蛋白质一级结构的氨基酸序列中隐含了蛋白质二级、三级结构的信息。而蛋白质二级结构预测又可为蛋白质三维结构预测和蛋白质功能预测提供重要信息。因此从一级氨基酸序列预测二级结构及蛋白质的性质是药物研发中的重要任务。表 1 简要列举了 DL 在蛋白质结构和功能预测中的应用。

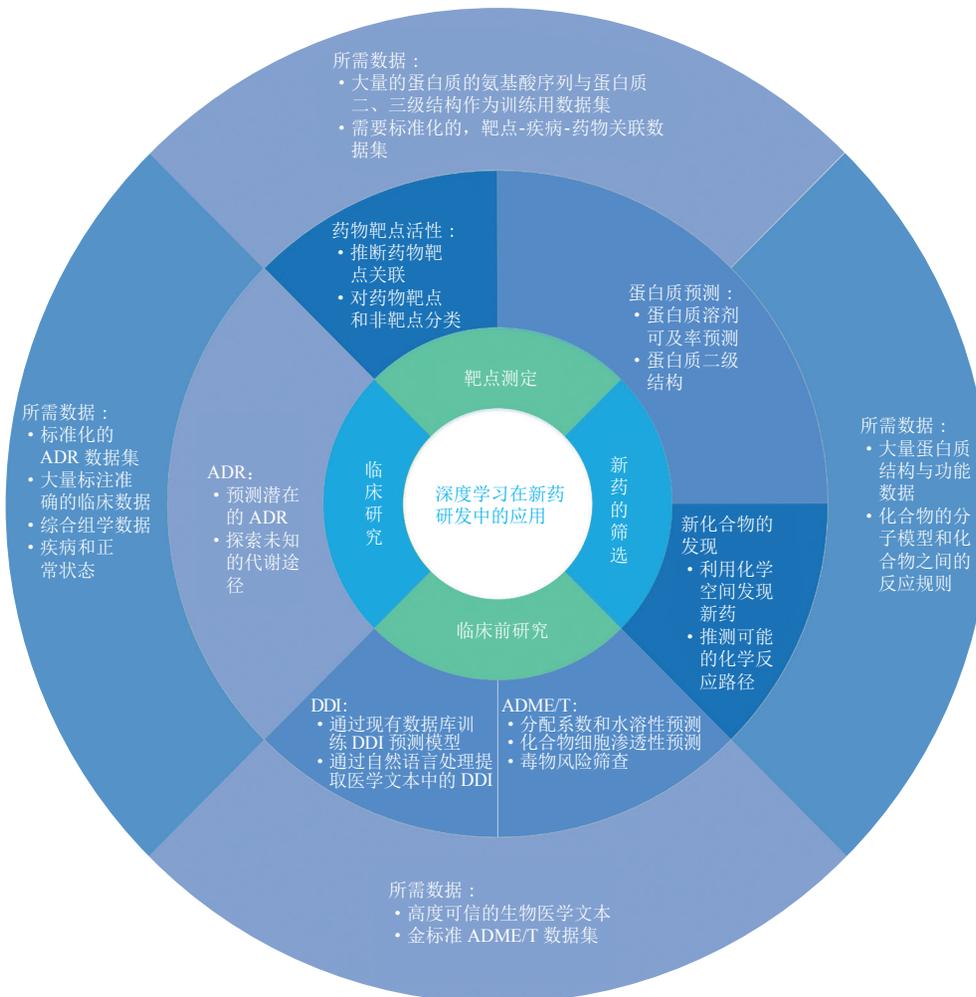
尽管近年来 X-射线晶体学和冷冻电镜技术的不断发展在蛋白质结构解析上获得突破, 但其检测蛋白质的成本过高, 利用 DL 对蛋白质进行预测显然是一个更高效的方法。通过对数据库提供的蛋白质数据特征提取, 预测出蛋白质结构与功能, 为解决蛋白质结构和功能的预测问题提供了可能的途径, 并在蛋白质结构和功能预测方面取得了较好的结果。

2.2 活性药物靶点的确定

药物靶点与疾病或生物分子的病理状态相关, 药物靶点的确定是药物研究和开发的基础。传统的药物发现主要遵循“一种药物, 一种靶点, 一种疾病”的观念, 最近越来越多的研究人员接受了药物靶点是多种靶蛋白的观点^[15-16], 并且多种靶蛋白倾向于出现在同一种疾病中^[17]。因此, 如何快速准确地识别药物与靶点之间复杂的相互作用已成为药物开发的关键。

Pu 等^[18]采用 CNN 训练检测和分类核苷酸与血红素结合位点, 准确度达到了 95%, 且实验模型能够推广到类固醇结合蛋白和肽酶。Hamanaka 等^[19]提出的 DL 模型在检测药物活性靶点时可以在在保证 98.2% 的准确率的情况下对 400 万个数据进行计算。Wen 等^[20]首先对未处理的原始数据进行预处理, 标记出已知的药物靶点相互作用, 然后应用已知的标记过的药物靶点对来训练分类模型, 该模型的 10-折交叉验证的曲线下面积 (area under curve,

AUC) 得分为 0.9158 ± 0.0059 , 该得分越趋近于 1 则说明效果越好, 该模型可进一步用于预测新靶点。



ADME/T: absorption, distribution, metabolism, excretion and toxicity (吸收、分布、代谢、排泄及药物毒性); ADR: adverse drug reaction (药物不良反应); DDI: drug-drug interaction (药物相互作用)

图 1 新药研发各环节中深度学习的应用

Figure 1 Application of DL in different steps of new drug research and development

表 1 深度学习在蛋白质结构和功能预测中的应用

Table 1 Application of DL in the prediction of protein structure and function

研发人员	深度学习方法	数据来源	预测内容	准确率	评价	参考文献
Khurana 等	CNN	蛋白质序列数据库: k-mer结构数据库	蛋白质溶解度	0.77	优于所有已知基于序列的溶解度预测方法	[11]
Fang等	深度初始内部网络 (Deep3I)		蛋白质结构		开发为软件工具MUFOLD-SS	[12]
Hou等	一维深度卷积神经网络 (DeepSF)	蛋白质结构分类数据库: SCOP (structural classification of protein)	蛋白质结构	SCOP1.75数据库准确率为 0.75; SCOP2.06数据库准确率为 0.73	对于序列突变、插入、缺失的预测有稳定性, 并且可以迁移应用到其他蛋白质模式识别问题, 例如蛋白质聚类, 比较和排序	[13]
Hanson等	长短期记忆网络 (long short-term memory, LSTM)	DM3000显微镜观察	预测无序蛋白		预测固有无序蛋白质和蛋白质的生物作用	[14]

Tian 等^[21]通过分层抽象学习药物靶点对的有用特征, 在平衡和不平衡数据集(平衡数据集是指各个样本数量差距不大, 而不平衡数据集则相反, 在一些模型中数据集是否平衡对预测结果有着不同影响)上的预测性能均比现有方法更好。Tsubaki 等^[22]结合化合物的图形神经网络(graph neural network, GNN)和蛋白质的 CNN 开发了新的复合蛋白相互作用预测技术。此外, 所提出的方法在不平衡数据集上明显优于现有方法。这表明由端到端 GNN 和 CNN 获得的化合物和蛋白质的数据驱动表示比从数据库获得的传统化学和生物学特征更稳健。Zong 等^[23]采用 DL 算法 DeepWalk 基于异构拓扑计算药物-药物和靶点-靶点的相似性, 基于“牵连犯罪”原则推断药物靶点关联, AUC 得分为 0.989 6。

Xie 等^[24]基于药物干扰和基因敲除试验收集的的药物和药物对, 靶点和靶点对的相似性数据库 L1 000 中的转录组数据提出了活性药物靶点的确定(drug target identification, DTI)预测框架, 训练集达到 98% 以上的准确率, 验证集准确率为 $90.53\% \pm 1.44\%$ 。同时使用其他 DTI 数据库如 STITCH、DGIdb 以及 CTD 验证了模型预测新 DTI 的能力, 均取得较好的结果。

计算机在处理复杂运算方面具有天然优势, 而 DL 可以进一步发现对象之间的隐性联系。DL 的方

法应用于复杂的药物与多靶点作用问题, 通过合适的网络模型得到潜在靶点, 将大大提高药物靶点研究的效率。

2.3 药物代谢动力学分析

药物代谢动力学(pharmacokinetics, PK)分析是药物研发过程的重要组成部分, 调查显示大约一半的候选药物由于 PK 性质或毒性不令人满意而无法进入市场^[25]。DL 正是 PK 分析的一项重要技术手段。

图 2 分别从药物的吸收、分布、代谢与毒性几个方面举例了 DL 应用。以 PK 的关键性质之一水溶性为例, 药物的水溶性将直接影响化合物在生物体中的吸收。Li 等^[26]建立了基于半监督学习模型的多层深度信念网络(deep belief network, DBN)来预测化合物的水溶性, 准确率达到 85.9%。不只是吸收分布的 PK 性质, 候选药物的代谢毒性也是其后续能否成药的关键因素。分析 1 824 种美国 FDA 批准的药物, 其中 29.6% 的药物经计算确定具有潜在的 hERG(human ether-à-go-go-related gene)抑制活性, 突出了 hERG 风险评估在早期药物发现中的重要性。为了在药物发现阶段和上市后监测中对 hERG 介导的心脏毒性进行风险评估, Cai 等^[27]开发了一种 DNN 模型用于预测药物发现和上市后监测过程中小分子的 hERG 阻滞剂。在验证集中, 最佳模型的 AUC 为 0.967。



图 2 深度学习在 PK 中的应用

Figure 2 Application of DL in PK

DL 为以传统模型为基础的 PK 分析带来了新的模型以及新的分析方式, 并且就特定的问题给出了更为精确的答案。

2.4 药物相互作用

当药物与另一种药物共同服用时, 此药物的预期功效可以发生显著改变。因此, 了解 DDI 对于减少意外的不良药物事件(adverse drug event,

ADE) 的发生以及在治疗疾病时产生最大化协同效益至关重要。另外由 DDI 引起的 ADR 是药物退出市场的主要原因之一^[28]。随着用于疾病治疗的多种药物(至少 2 种药物)的处方数量持续增加, 了解 DDI 的意义越来越大。探索用于检测相互作用药物的大量药物组合的最实用方法是通过计算机 DDI 检测。

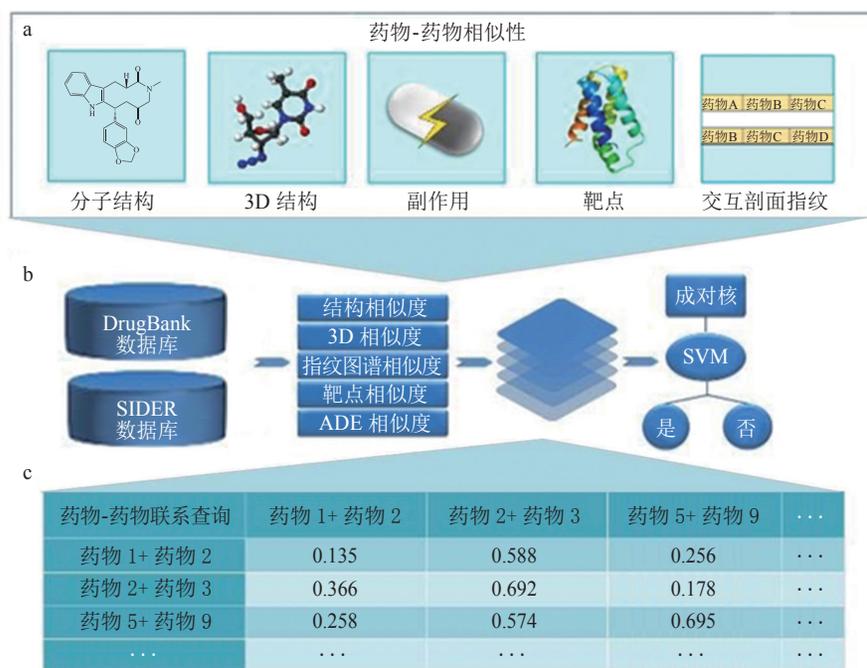
虽然一些已知的 DDI 可以在专门建立的数据库中 找到, 但大多数信息仍然埋藏在文献中。因此, 迫切需要从 生物医学文本中自动提取 DDI。为了 检验仅使用字嵌入作为输入特征的 CNN 是否可以 成功应用于生物医学文本中的 DDI 分类, Suárez-Paniagua 等^[29] 提出仅具有一个隐藏层的 CNN 架构, 使得模型在计算上更有效。Jari 等^[30] 使用 CNN 同时提取事件和关系, 与不同的向量空间嵌入一起 应用于各种文本分类任务。Zhao 等^[31] 提出语法卷积 神经网络 (syntax convolutional neural network, SCNN), 基于单词嵌入、语法单词嵌入来使用句 子的句法信息, 引入位置和词性特征以扩展每个单 词的嵌入, 引入自动编码器来编码传统的特征文本 词袋 (稀疏 0-1 向量) 作为全连接向量。

除 CNN 外 RNN 也常用于生物医学关系提取, Zhang 等^[32] 基于候选句子的依赖图生成最短依赖路 径 (shortest dependent path, SDP), 将 SDP 划分为 依赖词序列和关系序列。RNN 和 CNN 分别用于自动 学习句子序列和依赖序列的特征。最后, 将 RNN 和 CNN 的输出特征结合起来检测和提取生物医学关系。

Sahu 等^[33] 提出了 3 种长短时记忆模型 (long

short-term memory, LSTM) 网络模型, 即双向 长短时记忆网络 (bi-directional long short-term memory, Bi-LSTM), 基于 Attention 模型的 Bi-LSTM (attention based bi-directional long short-term memory, AB-LSTM) 和基于联合模型的 AB-LSTM (joint attention based bi-directional long short-term memory, joint AB-LSTM)。这 3 种模型都使用文 字和位置嵌入作为潜在特征。此外, 使用 Bi-LSTM 网络允许从整个句子中提取隐含特征。2 个模型 AB-LSTM 和 joint AB-LSTM 也在 Bi-LSTM 层输出 中应用注意池, 以便为特征分配权重。

Song 等^[34] 利用支持向量机模型 (support vector machine, SVM) 建立了一个机器学习模型。所建立 的相似性测度包括二维分子结构相似性、三维结构 相似度、相互作用指纹图谱相似性、靶标相似性和 ADE 相似性。根据所建立的 5 种相似性度量方法, 将已知有作用的药物和可能有作用的药物进行处理, 使结构以数据形式表示, 并将处理结果作为 SVM 的输入向量。SVM 模型建立的思路以及所用数据库 如图 3 所示。最终, 此 SVM 模型预测准确率达到 0.97, 远高于之前的 DDI 模型。



a: 建立相似性测度; b: 将多维度的数据一同输入到建立的 SVM 模型中; c: 得到相互作用结果

图 3 基于 SVM 的药物相互作用预测

Figure 3 Prediction of DDI on SVM

2.5 药物不良反应

ADR 是一个严重的问题, 即尽管给予常规剂量的药物, 但仍会出现不良反应。据估计, 住院患者中有超过 200 万例发生严重 ADR, 每年导致大

于 100 000 例患者死亡^[35]。因此识别或预测潜在的 ADR 显得尤为重要, 表 2 介绍了 4 个 DL 在 ADR 方面的应用, 并对各自的结果进行了简要评价。

表 2 深度学习在 ADR 中的应用

Table 2 Application of DL in the prediction of ADR

研发人员	技术手段	结果评价	参考文献
Liang等	利用基于生成随机网络和隐马尔可夫链的新型深度学习模型	识别了单核苷酸多态性及其与药物代谢、疗效和ADR的关系, 较传统的贝叶斯模型更具成本效益	[36]
Wang等	利用药物化学、生物和生物医学信息设计了DNN模型	确定了药物的潜在ADR并预测新药的可能ADR	[37]
Chen等	通过条件随机域、Bi-LSTM和基于词汇特征的Bi-LSTM 3种模型	对中文ADR报告进行了命名实体识别以及实体关系抽取研究, 将ADR相关实体及关系提取	[38]

DL 帮助科研人员从庞大复杂的 ADR 报告中筛选并识别了可能具有临床价值的 ADR, 辅助药物的应用并且可能会揭示未知的药物代谢途径。

3 结语与展望

人工智能通过分析海量的文献、专利和临床结果, 找出潜在的、被忽视的通路、蛋白和机制等与疾病的相关性, 从而提出新的可供测试的假说, 通过实验验证已经取得一定的成果, 并显著提高新药研发流程中某些阶段的效率。值得注意的是, DL 提供了一种新的方法来探索基因组变异与药物基因组学研究中的多种事件之间的复杂关联, 为全基因组关联分析的数据复杂性提供有效的解决方案。但是目前 DL 并不能直接预测一个化合物能否成为药物, DL 在发现药物研发新机制和新靶点上的突破仍然面临以下挑战:

1) 基于大数据的人工智能, 擅长的是对已有知识的挖掘、重新组织和分配, 为 DL 算法提供大量数据, 并且将需要解决的问题正确地呈现出来, 它们才有可能捕捉到人类无法捕捉到的规律, 在海量的数据中寻找已有知识的关联性。在新药研发过程中, 新药研发规则不明确, 数据不明晰甚至含有错误信息, 而且充满了高度不确定性等问题, 给以高质量标识数据集为基础的 DL 人工智能带来巨大的挑战;

2) DL 依赖于高质量、有标识的大数据集。例如: Santos 等^[39]统计了美国 FDA 批准的 1 578 个

药物总共的靶点数目是 667 个, 而 Ensembl 数据库标注的潜在药物靶点就有 4 479 个, 药物靶点数据库 (therapeutic target database, TTD) 含有 2 360 个可成为药物靶点的分子信息, 包括 388 个已有相应药物上市的、461 个处于临床试验阶段的以及其他正在研究的和已停止研究的药物的靶点信息, 由于数据库涵盖面不同, 信息来源亦不同, 导致有实验支持的信息和预测的信息混杂在一起, 药物和靶点的命名也未采用统一规则, 如何整合成 DL 依赖的靶点数据是关键;

3) DL 是非常出色的算法工具, 能够学习数据, 但无法判断数据准确性, 尤其是遇到不常见的情况时, 它缺乏灵活性, 表现并不好。同时存在无法修正学习结果, 除非重新训练的问题;

4) DL 的内部机制一直是困扰现今科学家的难题, DL 是一个“黑箱”, 药物在人体中作用的机制是另一个“黑箱”。通过 DL 研究药理问题被视为用一个“黑箱”代替另一个“黑箱”, 也就是说 DL 并没有实际解决药物机制这一重大难题。即 DL 仅展示了可能的结果, 而没找到真正的因果关系;

5) DL 的评价机制仍然存在欠缺。DL 具备发现隐藏在复杂的生物系统下的各种关系的能力, 帮助药物研发找到了一个模型来解释生物复杂系统中发生的事情, 但模型预测结果依然需要实验验证。如何用少量合适的实验使得药物研发人员进行有效验证和评价 DL 的结果是一个待解决的问题。

综上, 面对药物研发需要解决的问题的多样性,

也需要有更为灵活和细致入微的思考方式, 构建适合药物研发各阶段的特殊 DL 模型, 并且将这些模

型整合才能在未来实现智能的药物研发。

[参考文献]

- [1] Vohora D, Singh G. *Pharmaceutical Medicine and Translational Clinical Research*[M]. New York: Academic Press, 2017.
- [2] DiMasi J A, Grabowski H G, Hansen R W. Innovation in the pharmaceutical industry: new estimates of R&D costs[J/OL]. *J Health Econ*, 2016, 47: 20-33[2020-01-01]. <https://www.sciencedirect.com/science/article/pii/S0167629616000291>. Doi:10.1016/j.jhealeco.2016.01.012.
- [3] Burbidge R, Trotter M, Buxton B, et al. Drug design by machine learning: support vector machines for pharmaceutical data analysis[J]. *Comput Chem*, 2001, 26(1): 5-14.
- [4] Doucet J P, Barbault F, Xia H, et al. Nonlinear SVM approaches to QSPR/QSAR studies and drug design[J]. *Curr Comput Aided Drug Des*, 2007, 3(4): 263-289.
- [5] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural Comput*, 2006, 18(7): 1527-1554.
- [6] Niven R N. Back to Biology[CP/OL].[2020-01-01]. <https://www.berghealth.com/>.
- [7] Narain R N, Moser A J, Ramanathan R, et al. Project survival: interrogative biology platform mediated discovery of molecular markers for detection, stratification and outcomes in pancreatic cancer[J]. *Cancer Res*, 2016, 76(14 Suppl): 786-786.
- [8] IBM. IBM Cognos Analytics on Cloud[CP/OL].[2020-01-01]. <https://www.ibm.com/cn-zh/products/category/technology/cognitive-computing-and-AI>.
- [9] Lu T, Collins J, Li H, et al. Deciphering biological complexity to unlock drug discovery[CP/OL].[2020-01-01]. <https://www.enginebio.com/>.
- [10] Anfinsen C B. Principles that govern the folding of protein chains[J]. *Science*, 1973, 181(4096): 223-230.
- [11] Khurana S, Rawi R, Kunji K, et al. DeepSol: a deep learning framework for sequence-based protein solubility prediction[J]. *Bioinformatics*, 2018, 34(15): 2605-2613.
- [12] Fang C, Shang Y, Xu D, et al. MUFOLD-SS: new deep inception-inside-inception networks for protein secondary structure prediction[J]. *Proteins*, 2018, 86(5): 592-598.
- [13] Hou J, Adhikari B, Cheng J L, et al. DeepSF: deep convolutional neural network for mapping protein sequences to folds[J]. *Bioinformatics*, 2018, 34(8): 1295-1303.
- [14] Hanson J, Yang Y D, Paliwal K, et al. Improving protein disorder prediction by deep bidirectional long short-term memory recurrent neural networks[J]. *Bioinformatics*, 2017, 33(5): 685-692.
- [15] Hopkins A L. Network pharmacology: the next paradigm in drug discovery[J]. *Nat Chem Biol*, 2008, 4(11): 682-690.
- [16] Chen X, Ren B, Chen M, et al. NLLSS: predicting synergistic drug combinations based on semi-supervised learning[J]. *PLoS Comput Biol*, 2016, 12(7): e1004975[2020-01-01]. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004975>.
- [17] Yamanishi Y, Araki M, Gutteridge A, et al. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces[J]. *Bioinformatics*, 2008, 24(13): 232-240.
- [18] Pu L, Govindaraj R G, Lemoine J M, et al. DeepDrug3D: classification of ligand-binding pockets in proteins with a convolutional neural network[J/OL]. *PLoS Comput Biol*, 2019, 15(2): e1006718[2020-01-01]. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006718>.
- [19] Hamanaka M, Taneishi K, Iwata H, et al. CGBVS-DNN: prediction of compound-protein interactions based on deep learning[J/OL]. *Mol Inform*, 2017, 36(1/2): 1600045[2020-01-01]. <https://onlinelibrary.wiley.com/doi/full/10.1002/minf.201600045>.
- [20] Wen M, Zhang Z, Niu S, et al. Deep-learning-based drug-target interaction prediction[J]. *J Proteome Res*, 2017, 16(4): 1401-1409.
- [21] Tian K, Shao M Y, Wang Y, et al. Boosting compound-protein interaction prediction by deep learning[J/OL]. *Methods*, 2016, 110:64-72[2020-01-01]. <https://www.sciencedirect.com/science/article/pii/S1046202316301992>. Doi:10.1016/j.ymeth.2016.06.024
- [22] Tsubaki M, Tomii K, Sese J, et al. Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences[J]. *Bioinformatics*, 2019, 35(2): 309-318.
- [23] Zong N, Kim H, Ngo V, et al. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations[J]. *Bioinformatics*, 2017, 33(15): 2337-2344.
- [24] Xie L W, He S, Song X Y, et al. Deep learning-based transcriptome

- data classification for drug-target interaction prediction[J/OL]. *BMC Genomics*, 2018, 19(Suppl 7): 667[2020-01-01]. <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-018-5031-0>.
- [25] Waterbeemd H V D, Gifford E. ADMET in silico modelling: towards prediction paradise?[J]. *Nat Rev Drug Discov*, 2003, 2(3): 192-204.
- [26] Li H, Yu L, Tian S W, *et al*. Deep learning in pharmacy: the prediction of aqueous solubility based on deep belief network[J]. *Autom Control Com S*, 2017, 51(2): 97-107.
- [27] Cai C P, Guo P F, Zhou Y D, *et al*. Deep learning-based prediction of drug-induced cardiotoxicity[J]. *J Chem Inf Model*, 2019, 59(3): 1073-1084.
- [28] Onakpoya I J, Heneghan C J, Aronson J K, *et al*. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature[J/OL]. *BMC Med*, 2016, 14: 10[2020-01-01]. <https://bmcmecicine.biomedcentral.com/articles/10.1186/s12916-016-0553-2>.
- [29] Suárez-Paniagua V, Segura-Bedmar I, Martínez P, *et al*. Exploring convolutional neural networks for drug-drug interaction extraction[J/OL]. *Database(Oxford)*, 2017, 2017: bax019[2020-01-01]. <https://academic.oup.com/database/article/doi/10.1093/database/bax019/3854772>.
- [30] Jari B, Tapio S. Biomedical event extraction using convolutional neural networks and dependency parsing[C]. Melbourne: Proceedings of the BioNLP 2018 workshop, 2018.
- [31] Zhao Z H, Yang Z H, Luo L, *et al*. Drug drug interaction extraction from biomedical literature using syntax convolutional neural network[J]. *Bioinformatics*, 2016, 32(22): 3444-3453.
- [32] Zhang Y J, Lin H F, Yang Z H, *et al*. A hybrid model based on neural networks for biomedical relation extraction[J/OL]. *J Biomed Inform*, 2018, 81: 83-92[2020-01-01]. <https://www.sciencedirect.com/science/article/pii/S1532046418300534?via%3Dihub>. Doi: 10.1016/j.jbi.2018.03.011.
- [33] Sahu S K, Anand A. Drug-drug interaction extraction from biomedical texts using long short-term memory network[J/OL]. *J Biomed Inform*, 2018, 86: 15-24[2020-01-01]. <https://www.sciencedirect.com/science/article/pii/S1532046418301606?via%3Dihub>. Doi: 10.1016/j.jbi.2018.08.005.
- [34] Song D L, Chen Y, Min Q, *et al*. Similarity-based machine learning support vector machine predictor of drug-drug interactions with improved accuracies[J]. *J Clin Pharm Ther*, 2019, 44(2): 268-275.
- [35] World Health Organization. Safety of medicines: a guide to detecting and reporting adverse drug reactions: why health professionals need to take action[R]. Geneva: WHO, 2002.
- [36] Liang Z, Huang J X, Zeng X, *et al*. DL-ADR: a novel deep learning model for classifying genomic variants into adverse drug reactions[J/OL]. *BMC Med Genomics*, 2016, 9 Suppl 2:48[2020-01-01]. <https://bmcmecgenomics.biomedcentral.com/articles/10.1186/s12920-016-0207-4>.
- [37] Wang C S, Lin P J, Cheng C L, *et al*. Detecting potential adverse drug reactions using a deep neural network model.[J/OL]. *J Med Internet Res*, 2019, 21(2): e11016[2020-01-01]. <https://www.jmir.org/2019/2/e11016/>. Doi: 10.2196/11016.
- [38] Chen Y, Zhou C J, Li T X, *et al*. Named entity recognition from Chinese adverse drug event reports with lexical feature based BiLSTM-CRF and tri-training[J/OL]. *J Biomed Inform*, 2019, 96: 103252[2020-01-01]. [https://linkinghub.elsevier.com/retrieve/pii/S1532-0464\(19\)30171-6](https://linkinghub.elsevier.com/retrieve/pii/S1532-0464(19)30171-6). Doi: 10.1016/j.jbi.2019.103252.
- [39] Santos R, Ursu O, Gaulton A, *et al*. A comprehensive map of molecular drug targets[J]. *Nat Rev Drug Discov*, 2016, 16(1):19-34.



【专家介绍】尚靖: 教授, 博士生导师。先后就读兰州大学生物化学专业、新疆大学药物资源专业, 获德国柏林自由大学分子生物学专业理学博士学位。兼任中国药学会应用药理专业委员会副主任委员、中国药理学学会制药工业委员会副主任委员兼秘书长, 江苏省动物协会实验动物专业委员会的主任委员。作为课题负责人荣获 2003、2005、2013 年度省级科技进步二等奖, 曾获新疆维吾尔自治区人民政府颁发的“青年科技奖”, 教育部“新世纪优秀人才支持计划”, 江苏省“六大人才高峰”计划, 中科院“百人计划”。目前是中国药科大学“皮肤健康及疾病干预”研究平台负责人, 生物与医药专业化化妆品与皮肤健康方向研究生专业负责人。

先后连续主持“九五”、“十五”、“十一五”、“十二五”、“十三五”5 项国家重大新药创制专项的新药开发项目; 主持国家自然科学基金 3 项。截至目前, 出版专著 2 部, 其中英文 1 部, 发表论文 161 余篇, 其中 SCI 收录论文 71 篇; 作为第一申请人申请专利 20 余项, 授权国内专利 8 项, 申请国际专利 11 项, PCT 专利 4 项, 授权国际专利 6 项。主持完成 3 个中药、天然药 5 类的创新药临床前研究。作为主要完成人研制的驱虫斑鸠菊注射液(国药准字 Z20063652)已实现产值 4.68 亿元。