

人工智能在药物再定位的应用

陈杨¹, 陈轶伦², 曾文杰², 廖俊^{2*}

(中国药科大学基础医学与临床药学院¹, 理学院², 江苏 南京 211198)

[摘要] 药物再定位是利用已上市药物或临床研究失败的药物等进行药物新适应证探索的一种方法, 因其通过缩短研发周期、降低研发成本可解决目前药物研发面临的漫长周期和高昂费用问题, 而获得了医药行业的广泛认可。随着医药大数据、计算能力和深度学习算法的巨大进步, 人工智能运用于药物再定位在时间和成本方面有明显优势。通过对医药大数据进行采集、加工和处理, 并进一步分析和挖掘, 可以有效建立起药物-疾病之间的联系, 推动药物再定位的发展。本研究将从药物、疾病以及真实世界数据 3 个方面对人工智能参与药物再定位的方法进行综述, 并进一步探讨其中可能面临的挑战, 以充分理解人工智能在药物再定位中的应用, 推动人工智能与药物研发融合发展。

[关键词] 药物再定位; 人工智能; 真实世界研究; 真实世界数据

[中图分类号] R95; TP39 **[文献标志码]** A **[文章编号]** 1009-0959(2022)04-0334-08

Application of Artificial Intelligence in Drug Repurposing

CHEN Yang¹, CHEN Yilun², ZENG Wenjie², LIAO Jun^{2*}

(School of Basic Medicine and Clinical Pharmacy¹, School of Science², China Pharmaceutical University, Jiangsu Nanjing 211198, China)

[Abstract] Drug repurposing is a method for exploring new indications for marketed drugs or drugs that have failed clinical studies, and has been widely recognized by the pharmaceutical industry because it can solve the risk of long cycle and high costs currently faced in drug research and development (R&D) by shortening the R&D cycle and reducing R&D costs. With the tremendous progress of pharmaceutical big data, computing power and deep learning algorithms, the application of artificial intelligence (AI) in drug repurposing has obvious advantages in time and cost. Through the collection, processing and handling of pharmaceutical big data, and further analysis and mining, the connection between drugs and diseases can be effectively established and the development of drug repurposing can be promoted. This review will elaborate the methods of AI participating in drug repurposing from three aspects of drugs, diseases, and real world data, and further discuss the possible challenges, so as to fully understand the application of AI in drug repurposing, and promote the integrated development of AI and drug R&D.

[Key Words] Drug repurposing; Artificial intelligence; Real world study; Real world data

药物研发是一个复杂的过程, 面临着耗时长、投入高、风险大的挑战, 需要涉及多个重要的环节和阶段, 一款创新药物从研发到最终上市通常需要 10 年甚至更长的时间, 研发投入的成本有的甚至超过了 28 亿美元^[1,2]。药物研发成本在近 20 年显著增加, 一个重要原因是大量的化合物需要在临床前进行测试, 并且其中很大一部分化合物由于临床效果不佳、毒性较强等原因而付诸东流, 造成了研发成本的高损耗^[1,3]。

针对新药开发的这些挑战, 一个有效的应对方法是利用现有药物开发新的适应证, 这一方法已经在西地那非、米诺地尔等多个药物中得到了充分的应用^[4,5]。已批准上市的药物在前期研究中通过了临床试验的全部阶段, 因此其对人体的安全性是已知和公认的, 如果对已获批药物提出新的适应证, 该药物将有机会跳过临床前研究而直接进入临床试验, 从而大大减少研发时间和研发成本, 这对学术界和工业界的研究人员来说是一项非常重要的研发策略^[6]。

这种将老药重新定位以用于治疗其他疾病的方法称为药物再定位 (drug repurposing/drug repositioning)。Ashburn 和 Thor 在 2004 年提出“药物再定位”这一具有里程碑意义的概念, 并详细阐述了药物再定

[作者简介] 陈杨, 男, 在读硕士研究生, 研究方向: 医药大数据与人工智能。
E-mail: chen yang2121@foxmail.com

***[通讯作者]** 廖俊, 男, 博士, 副教授, 研究方向: 医药大数据与人工智能。
E-mail: liaojun@cpu.edu.cn

位在识别和发现有药物新用途方面的作用^[7]。随着药物研发的不断深入,药物再定位的概念也有了新的变化。再定位药物不再局限于已批准上市的药物,而是扩大到临床研究中失败的药物或尚待开

发的临床候选药物^[8]。如表1列举了部分通过药物再定位获得新适应证的药物,其中大部分都是基于现有药物进行再定位或者通过临床研究发现进行药物再定位。

表1 部分药物通过药物再定位获得新适应证

药物名称	原适应证	新适应证	批准时间	再定位的药物类型
齐多夫定	癌症	HIV	1987年	基于临床研究失败的药物
米诺地尔	高血压	脱发	1988年	基于现有药物
西地那非	心绞痛	勃起功能障碍/肺动脉高压	1998年/2005年	基于临床研究失败的药物
沙利度胺	孕吐	麻风性结节性红斑/多发性骨髓瘤	1998年/2006年	基于撤市药物
度洛西汀	抑郁症	压力性尿失禁	2004年	基于现有药物
利妥昔单抗	非霍奇金淋巴瘤	类风湿性关节炎	2006年	基于现有药物
雷洛昔芬	骨质疏松症	乳腺癌	2007年	基于现有药物
芬戈莫德	免疫抑制剂	多发性硬化症	2010年	基于临床研究的候选药物
达泊西汀	抑郁症	早泄	2012年	基于临床研究的候选药物
托吡酯	癫痫	肥胖	2012年	基于现有药物
富马酸二甲酯	抑菌防霉(工业)	多发性硬化症	2013年	基于临床研究的候选药物
屈昔多巴	帕金森综合征	神经源性体位性低血压	2014年	基于现有药物
地夫可特	糖皮质激素	杜氏肌营养不良症	2017年	基于现有药物
依达拉奉	自由基清除剂	肌萎缩侧索硬化症	2017年	基于现有药物
氨磺必利	精神分裂症	术后呕吐	2020年	基于现有药物
左旋酮康唑	类固醇生成抑制剂	库欣综合征	2021年	基于临床研究的候选药物

目前越来越多的化合物通过药物再定位的方式获得了新的适应证,这得益于人工智能和大数据的蓬勃发展。生物信息数据的爆炸式增长为药物再定位带来了更多的机遇,面对海量数据,研究人员可以利用人工智能进行大数据分析,发现某一疾病的治疗药物之间的分子相似性,并构建模型来预测哪些化合物拥有这种相似性;可以通过药物靶点相互作用分析药物的潜在靶点活性;可以利用高通量筛选技术快速测试药理活性^[6]。人工智能技术的应用加快了药物再定位的速度,研究人员不必再被动等待偶然机遇的出现,而是可以主动利用现有临床数据和生物信息数据等资源进行筛选和分析,为药物研发注入新的活力。

近年来,人工智能在药物再定位方面的相关研究呈现明显的上升趋势,如图1是PubMed数据库中人工智能与药物再定位相关文献的发表情况,通过文献计量学分析发现,以机器学习和深度学习为主的人工智能技术从2013年左右开始逐步应用于药物再定位研究,并且机器学习比深度学习应用更广泛。此外,人工智能辅助药物再定位研究在2020—2021年间有较大幅度的增长,这可能与新

冠肺炎(COVID-19)疫情突发以及人工智能技术的发展有一定的关系。新冠肺炎疫情的大流行给全球健康带来了严峻压力,各国面临“无药可用”的局面,为了尽快研发药物,遏制疫情蔓延,研究人员对新冠病毒潜在治疗靶点进行了全面剖析,并利用人工智能技术对一系列具有抗病毒活性的临床药物或化合物进行了虚拟筛选,寻找并开发具有抗新冠病毒活性的药物^[9]。药物再定位作为常规药物开发的一种有效补充,在降低研发成本和加速药物研发方面已经展现出了明显的优势,人工智能的参与可以最大限度减少研发的失败率,进一步巩固药物再定位的优势^[10]。本研究基于这些背景进一步探讨人工智能在药物再定位方面的技术应用,体现人工智能赋能药物研发的重要价值和意义。

1 药物再定位研究

与常规药物开发相比,利用药物再定位方法开发新适应证的药物在研发周期和研发流程方面有较大幅度的精简,人工智能技术的参与将进一步缩短研发周期,降低研发成本(见图2)。通常,药物再定位主要包含以下步骤^[11]:①候选化合物识别和确证;②临

床前研究;③Ⅱ期和Ⅲ期临床试验疗效和安全性评估;④药品注册和上市。其中确定具有合适适应证的候选化合物至关重要。与常规药物开发不同,药物再定位并不是从无到有的过程,而是建立在以往的各种研究基础之上,甚至包含着一定的偶然机遇。人们利用偶然发现或者已有数据资料提出假设并进一步研究,这种“假设-验证”的过程促进了现代药物研发之路的新发展。



图1 近10年来人工智能与药物再定位相关研究趋势图

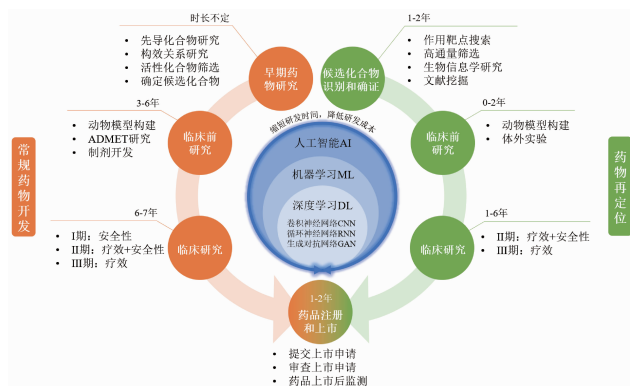


图2 常规药物开发与药物再定位的研发流程

在进行候选化合物识别和确证的过程中,人们往往会对具有相似性质的药物或疾病产生兴趣,认为拥有相似化学结构的药物可能具有相似的药理作用,拥有相似病理特征的疾病或许可以使用同一药物进行治疗,这种建立在相似性方面的假设为正确筛选化合物提供了一定的理论基础,也成为药物再定位的重要研究策略。“相似性假设”原理最早由 Johnson 和 Maggiora 提出,其核心思想是结构相似的化合物拥有相似的理化性质和生物活性^[12]。这一原理为药物的虚拟筛选和化合物的识别确证奠定了基础,药物再定位也在这个基础上进行了进一步的利用和拓展。

目前,药物再定位主要基于两个方向进行:①基于药物的相似性,即通过药物的分子活性相似性、化学结构相似性或不良反应相似性进行特征匹配,寻找合适的药物-靶标并模拟分子对接,预测药

物的新适应证;②基于疾病的相似性,即通过疾病的病理学特征、临床特征、疾病表型特征等构建疾病相似性网络,并与药物治疗谱和疾病基因谱共同创建药物-疾病-基因网络,推测和寻找可能具有新适应证的药物^[13]。

近年来,随着医药大数据和人工智能的发展,越来越多的研究开始追求证据的有效性、真实性和可溯源性,真实世界研究(real world study, RWS)这一概念应运而生^[14]。真实世界研究是通过收集真实世界环境中与患者有关的数据[真实世界数据(real world data, RWD)],通过分析,获得医疗产品的使用价值及潜在获益或风险的临床证据(真实世界证据)^[15]。真实世界数据来源广泛,包括大规模的临床试验数据、电子健康记录数据、药品上市后监测数据、疾病注册登记数据、医疗索赔数据等多样化数据,提供了许多药物-药物或药物-疾病之间关系的相关信息。研究人员可以利用人工智能框架对真实世界数据进行回顾性分析并构建模型,寻找和测试候选药物用于药物再定位^[16]。

由此可见,基于药物和疾病的相似性假设和大数据分析进行人工智能辅助药物再定位研究,进一步明确了药物-疾病-基因之间的关联,有助于药物潜在新适应证的开发,将对药物研发带来重要的帮助和影响。

2 人工智能与药物再定位结合研究方法

药物再定位需要从以往的研究中发现新的规律和性质。这些研究产生了大量药物和疾病相关的数据,人工智能可以借助这些数据驱动算法,从而完成模型构建,进行候选药物的快速筛选(见图3)。在人工智能的辅助下,基于药物相似性进行特征匹配和分子对接,可以进一步了解药物-靶标的相互作用,寻找具有脱靶效应的药物;基于疾病基因和疾病病理搭建关联性网络,进行疾病相似性研究,可以帮助寻找具有相似治疗效果的药物;基于临床或药物上市后产生的真实世界数据进行大数据分析,为药物再定位探索了新的证据来源。

2.1 基于药物相似性的药物再定位与人工智能方法研究

2.1.1 特征匹配

特征匹配(signature matching)是基于一种药物的独特特征与另一种药物的特征进行比较^[17]。药物特征可以来源于分子活性、化学结构或不良反应等。通常分子活性或化学结构特征可用于药物-药物之间

的比较,即比较不同药物之间的相似性,寻找共同的作用机制,帮助识别现有药物的替代靶点,并揭示可用于临床应用的潜在脱靶效应^[17, 18]。不良反应特征是药物具有代表性的一种表型,导致相同不良反应的两种药物可能作用于共同的靶标或有相似的作用途径,特定不良反应的表型可能与疾病表型相似,这表明潜在的表达途径或生理系统可能受到药物和疾病的共同作用^[13]。

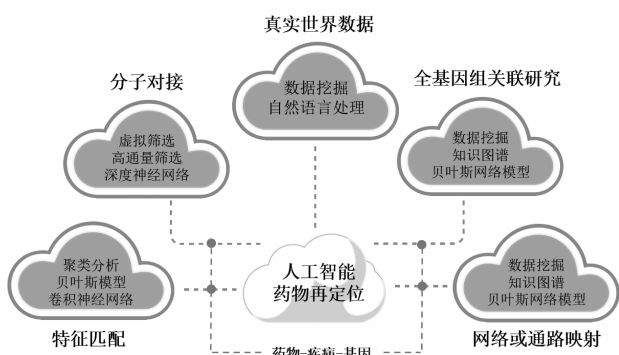


图3 人工智能与药物再定位结合研究方法

预测药物分子活性是进行药物再定位的重要一步,即预测药物分子对靶标的结合亲和力,分析药效团或定量构效关系(QSAR),研究小分子与生物大分子之间的相互作用和生理效应。Feinberg等^[19]提出神经网络模型PotentialNet,该模型用于评价蛋白-配体结合亲和力,通过分子特征预测助力药物发现。Hochuli等^[20]利用神经网络对蛋白质-配体复合物的结合姿势和亲和力预测评分并进行可视化,实现有效的虚拟筛选。Issa等^[21]报告了一种称为RepurposeVS的虚拟筛选方法,该方法可使用X射线蛋白质晶体结构可靠地预测药物-靶标特征,可用于预测药物再定位中的药物-靶标的关联,探索当前批准药物整体潜在的药物再定位空间。

有研究综合采用化学相似性方法进行药物再定位,该方法结合了化合物之间的结构相似性以及已建立的化合物-靶标关系的知识^[18]。研究人员将3 665种美国FDA批准的研究药物与数百个靶标进行了比较,并通过其配体定义了每个靶标,再使用相似度集成方法(SEA)将每种药物与每个靶标的配体集的2D结构相似性量化为期望值(E值)进行脱靶预测,最后通过实验测试以及与朴素贝叶斯分类器的预测进行了比较,确认了23个新的药物靶标关联,这对药物新适应证的开发起到了推动作用。

雷诺现象(Raynaud's phenomenon, RP)是血管对寒冷、潮湿或情绪压力的过度反应,通常表现为手指末端皮肤颜色的突然变化,出现指端缺血现象^[22]。一线治疗RP常使用钙通道阻滞剂或磷酸二酯酶-5抑制剂等血管扩张剂,但疗效并不理想^[23]。Zaza等^[24]研究人员利用WHO药物警戒数据库中的药品不良反应进行特征匹配,寻找可能对继发于系统性硬化症的RP有治疗效果的药物,即对现有无治疗RP适应证的药物进行药物再定位,寻找潜在的候选药物,发现使用血管扩张剂可能会诱发红斑性肢痛这一不良反应,这与RP的病理机制相反,因此对数据库中存在这一不良反应的药物进行筛选和提取,最后进行聚类分析,结果表明,阿仑单抗和富马酸可能具有治疗RP的潜在适应证。

2.1.2 分子对接

分子对接(molecular docking)是一种基于结构的计算策略,用于分子水平上预测配体-靶标的相互作用^[25]。如果知道某一疾病涉及的受体靶标,那么可以针对该靶标寻找多种化合物配体;相反,分子对接也可以应用于反向筛选,根据结构互补性识别已知配体的新型靶标^[17, 26]。Dakshanamurthy等^[27]利用一种称为TMFS(Train-Match-Fit-Streamline)的高通量算法对2 335种人类蛋白质晶体结构与3 671种美国FDA批准的药物进行了广泛的分子拟合计算,发现甲苯咪唑(抗寄生虫药物)具有抑制血管内皮生长因子受体-2(VEGFR-2)的结构潜力,可用于抑制癌症扩散。Zahler等^[28]描述了一种虚拟反向筛选方法,以识别蛋白激酶抑制剂靛玉红的衍生物的新靶标;他们对来自蛋白质数据库的大量蛋白质结构进行了虚拟测试并与抑制剂进行结合,发现磷酸肌醇依赖性蛋白激酶1(PDK1)是靛玉红衍生物6BIO的一个靶点目标。

2.2 基于疾病相似性的药物再定位与人工智能方法研究

2.2.1 网络或通路映射

基于网络或通路映射(network or pathway mapping)的方法广泛用于识别可能具有再定位潜力的药物或药物靶点,将基因组信息、病理生理学信息、蛋白质相互作用信息等数据构建为大型网络,从不同路径的大型网络中,可以识别由几个特定目标组成的特定网络,还可以利用通路提供相关靶标上游或下游基因的信息,帮助识别用于再定位的候选药物^[17, 29, 30]。

Mejia-Pedroza等^[31]基于通路对乳腺癌分子分

型的治疗进行了药物再定位。他们获得了 METABRIC 数据库和 TCGA 数据库的相关样本数据集,然后根据 PAM50 算法将样本分类为各自的乳腺癌分子分型,再通过 Pathifier 算法确定了每个乳腺癌分型中最失调的通路。根据药理学数据库 (DrugDB) 对每个失调通路确定了药理学靶标基因以及药物-靶标相互作用的性质,最后通过差异表达分析,寻找可进行药理学调节且当前未用于相应乳腺癌分型治疗的失调通路。

Rivero-Garcia 等^[32]通过网络分析将药物与疾病模块联系起来,发现一种药物平均针对 4 个疾病模块,但模块之间的相似性普遍较低,而通过药物再定位拓展邻近的药物目标时,针对的疾病模块有了很大的增加。其研究发现,药物再定位更多地依赖于疾病之间共享的靶蛋白,而不是药物的多药理学特性。研究者将药物再定位和网络模块分析应用于新冠肺炎患者,结果发现福坦替尼 (Fostamatinib) 是模块覆盖率最高的药物。

2.2.2 全基因组关联研究

全基因组关联研究 (genome-wide association studies, GWAS) 旨在识别与常见疾病相关的遗传变异,从而提供对疾病生物学的见解,获得的数据也可能有助于确定新的目标,其中一些可以在药物治疗的疾病和 GWAS 研究的疾病表型之间共享,从而进行药物再定位^[33]。在人工智能发展时代,机器学习和深度学习的使用正在改变遗传分析、药物开发和再定位的生物网络构建,还可以通过多基因风险评估和生物文献的自然语言处理分析进行疾病风险预测^[34]。

Fang 等^[35]开发了一种基于网络的人工智能框架,该框架能够将多组学数据与人类蛋白质-蛋白质相互作用组网络相结合,以准确推断由 GWAS 识别的基因变体影响的准确药物靶点,从而识别阿尔兹海默病 (Alzheimer's disease, AD) 新疗法,确定 AD 药物再定位的候选目标。具体来说,研究人员首先使用贝叶斯算法从 GWAS 的发现中预测 AD 风险基因 (ARG) 作为潜在的药物靶标,其次通过在人类蛋白质-蛋白质相互作用组网络模型下量化药物靶标和 ARGs 之间的网络邻近度得分来优先考虑候选药物。

Wang 等^[36]提出了一种药物再定位系统 Geno-Predict,用来寻找治疗前列腺癌 (prostate cancer, PC) 的候选药物。该系统利用 GWAS 生成的综合疾病遗传数据和大规模药物-疾病治疗关系知识库间

接推断再定位的候选药物,系统地将治疗药物从一种疾病重新定位到另一种疾病。研究人员首先构建了一个由 882 个节点和 200 758 条联系组成的遗传疾病网络 (GDN),并应用基于网络的排序算法从 GDN 中查找与 PC 遗传相关的疾病,并使用药物优先算法来重新定位 PC 遗传相关疾病中的药物以治疗 PC。

2.3 基于 RWD 的药物再定位与人工智能方法研究

使用人工智能进行药物再定位的另一个重要方面是充分利用 RWD 来寻找有效的候选化合物。与随机对照试验 (randomized controlled trial, RCT) 这类具有严格筛选条件和标准干预措施的“金标准”不同,通过 RWS 得到的 RWD 样本量更大、研究环境更真实、人群更为多样化,能够为治疗开发、结果研究、患者护理、医疗保健系统研究、质量改进、安全监测和控制良好的有效性研究等提供相关信息^[14, 37]。电子健康记录 (electronic health records, EHR) 是 RWD 中具有代表性的一类数据,EHR 是以数字化形式存储的患者临床数据,包含大量结构化和非结构化数据,例如人口统计学信息、疾病诊断信息、药物处方信息、实验室检查结果、患者症状和体征的临床描述以及成像数据等,可用于药物发现、药物再定位和药物不良事件预测^[38]。

Xu 等^[39]利用范德比尔特大学医学中心和梅奥诊所两个先进的 EHR 数据库进行了一项回顾性队列研究,以评估二甲双胍 (降糖药物) 与癌症病例中总死亡率之间的关系。在这项研究中,研究人员利用自然语言处理技术从 EHR 非结构化的文本中提取出药物名称和特征信息等标准的临床信息,并利用这些信息与其他结构化信息构建风险模型,验证二甲双胍是否具有参与癌症化疗的潜力,结果表明二甲双胍与癌症死亡率降低有关。

Ribelles 等^[40]利用机器学习和自然语言处理技术开发了一个预测模型,该模型基于真实世界中激素受体阳性 (HR +)/人表皮生长因子受体-2 阴性 (HER2-) 的晚期乳腺癌患者的 EHR 数据,用于预测早期和晚期进展为 HR +/HER2-转移性乳腺癌的一线治疗。

除了 EHR 这类规范化管理的电子数据外,在线收集患者报告的数据也是 RWD 的有效补充,可以作为药物再定位的潜在数据来源。Frost 等^[41]利用 PatientsLikeMe 平台中患者报告的用药数据作为标签外处方的证据来源。PatientLikeMe 是一个面向患者、研

研究人员和护理人员的工具,可帮助用户做出治疗决策,以便管理症状和改善治疗结果。在这项研究中,研究人员探索了以患者为中心的在线研究平台如何通过有效汇总结构化的患者报告数据来补充传统试验,从而更深入地了解上市后药物,为监测药物安全性和有效性以及发现药物潜在新适应证提供了新的证据来源。

3 人工智能应用于药物再定位所面临的挑战

3.1 人工智能模型的“黑箱效应”

人工智能依赖于海量数据和具有大量参数的复杂模型结构进行智能决策,可以帮助诊断和预测疾病结果^[42],在医药领域发挥了巨大作用。然而,人工智能的这种复杂性使得人们难以准确理解和判断模型是如何进行决定或预测的,这种“黑箱效应”带来的不可解释性可能会引起一定的社会偏见,即使模型最终预测的结果趋于准确,仍然会带来一些质疑^[43]。人工智能应当尽可能提高透明性、可解释性、可靠性和可控性,并逐步实现可监督性和可溯源性。人工智能模型是基于数据和算法运行的,保证数据的可靠性和算法的可解释性,推动结果可视化,可以在一定程度上提高人们对人工智能预测的信任度。

尽管如此,“黑箱效应”的存在也让研究人员有了更多的创新空间,可以通过不断优化参数设置、部署合适的激活函数和设计优化算法来不断改进现有深度学习方法。研究人员将不断优化的深度学习方法应用于新冠肺炎患者的药物再定位,并扩展到其他复杂疾病,为疾病提供适当的治疗策略,寻找潜在的治疗药物^[44]。

3.2 数据可用性

药物再定位过程中需要广泛用到数据,数据也是人工智能的基础,除了生物信息数据库和一些医疗数据库等公共数据库外,还有许多数据库无法被研究人员利用,这类数据库涉及到用户权限、共享限制、专利保护和数据机密等障碍,在数据保护期内无法被其他研究人员利用。因此应当建立有效的合作共享机制,提高数据的利用度,充分发挥数据在药物研发方面的作用。

3.3 数据质量

人工智能在药物发现和临床决策方面提供有效结果的准确性在很大程度上取决于数据质量。然而 RWD 来源广泛,数据之间的异质性较大,例如 EHR 中的非结构化数据由于不同医生记录的方式不同,其中造成的差异就会较大,导致难以提取到

统一形式的数据库。除了数据异质性外,数据缺失、大量重复等问题也会对数据质量造成影响,因此需要建立适用于药物研发和再定位研究的更高质量的药物适应证数据库,保证数据能以规范的数据标准进行记录^[45]。

3.4 数据伦理性

药物再定位过程需要收集和共享大量医学数据,这些数据可用于人工智能的模型构建,提高药物再定位的速度,但数据中包含的患者信息可能会引发公众对患者隐私的担忧^[46]。在这种情况下,需要构建合理可信的隐私保护框架,促进数据在框架支持的范围内安全流动,同时可以通过数据脱敏将患者个人信息隐藏,充分保障患者的隐私^[47]。对于患者隐私性较强的数据,需要提前向伦理委员会进行备案,避免进行违背伦理的科研行为。

4 总结与展望

人工智能发展 60 余年,已经在医学治疗和药物研发方面积累了大量的经验^[48,49]。当前的大数据时代,海量的医药数据可以成为药物研发的重要武器,研究人员不必依靠偶然机遇去获得药物再定位的线索,而是可以基于这些数据主动利用人工智能技术去寻找和识别线索,并通过计算模型进行药物再定位的模拟预测,进一步降低研发风险,提高研发速度,减少研发成本。借助药物再定位的策略,人工智能在特征匹配和分子对接以及网络或通路映射和全基因组关联研究方面发挥着至关重要的作用,从药物和疾病两方面对药物新适应证的开发提供了参考和帮助。除了传统角度,近年来研究火热的 RWD 也在药物再定位过程中发挥着日益显著的作用。利用人工智能辅助处理分析 EHR 这类真实的患者健康数据,可以进一步明确药物-疾病-基因之间的关系,为重新定位药物适应证助力。

毋庸置疑,数据的真实性、可靠性和完整性是成功保障药物再定位研究有效开展的重要基石,也是人工智能成功应用的基础。尽管人工智能与药物再定位研究在数据质量、数据可用性以及数据伦理性等方面面临着挑战,但并非不可克服。通过强化数据的规范化管理,促进数据在安全私密的框架内合理共享,将有利于药物再定位的科学开展。研究过程中除了要关注药物和疾病本身,也要特别注重监管科学问题,要建立长期有效的机制保障数据能够被合理合法应用,激励人工智能充分参与到药物再定位过程中来。

人工智能辅助药物再定位的研发前景广阔,随着

医药大数据的进一步丰富和发展,相信未来将会有越来越多的新适应证药物在罕见病、传染病等重大疾病领域造福于患者。

参考文献

- [1] Ekins S, Puhl AC, Zorn KM, et al. Exploiting machine learning for end-to-end drug discovery and development [J]. *Nat Mater*, 2019, 18 (5): 435-441.
- [2] Grabowski HG, Moe JL. Impact of economic, regulatory, and patent policies on innovation in cancer chemoprevention [J]. *Cancer Prev Res (Phila)*, 2008, 1(2):84-90.
- [3] Fleming N. How artificial intelligence is changing drug discovery[J]. *Nature*, 2018, 557(7707):S55-S57.
- [4] Ghofrani HA, Osterloh IH, Grimminger F. Sildenafil:from angina to erectile dysfunction to pulmonary hypertension and beyond[J]. *Nat Rev Drug Discov*, 2006, 5(8):689-702.
- [5] Suchonwanit P, Thammarucha S, Leerunyakul K. Minoxidil and its use in hair disorders:a review[J]. *Drug Des Devel Ther*, 2019, 13:2777-2786.
- [6] Nosengo N. Can you teach old drugs new tricks? [J]. *Nature*, 2016, 534(7607):314-316.
- [7] Ashburn TT, Thor KB. Drug repositioning:identifying and developing new uses for existing drugs[J]. *Nat Rev Drug Discov*, 2004, 3(8):673-683.
- [8] Allarakhia M. Open-source approaches for the repurposing of existing or failed candidate drugs:learning from and applying the lessons across diseases [J]. *Drug Des Devel Ther*, 2013, 7:753-766.
- [9] Wu C, Liu Y, Yang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods[J]. *Acta Pharm Sin B*, 2020, 10(5):766-788.
- [10] Mohanty S, Rashid M, Mridul M, et al. Application of artificial intelligence in COVID-19 drug repurposing[J]. *Diabetes Metab Syndr*, 2020, 14(5): 1027-1031.
- [11] Roessler HI, Knoers N, van Haelst MM, et al. Drug repurposing for rare diseases[J]. *Trends Pharmacol Sci*, 2021, 42(4):255-267.
- [12] Johnson MA, Maggiora GM. Similarity in chemistry. (Book Reviews; concepts and applications of molecular similarity) [J]. *Science*, 1991; 252. DOI:10.1126/science.252.5009.1189.
- [13] Dudley JT, Deshpande T, Butte AJ. Exploiting drug-disease relationships for computational drug repositioning [J]. *Brief Bioinform*, 2011, 12(4): 303-311.
- [14] 秦雪妮, 陈维生, 邵华, 等. 真实世界研究在医药领域的应用及研究方法[J]. *药学进展*, 2021, 45(7):512-523.
- [15] 国家药品监督管理局. 国家药监局药审中心关于发布《真实世界研究支持儿童药物研发与审评的技术指导原则(试行)》的通告(2020年第22号) [EB/OL]. (2020-08-27) [2022-03-14]. <https://www.nmpa.gov.cn/xxgk/ggtg/qtggtg/20200901104448101.html>.
- [16] Liu RQ, Wei L, Zhang P. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data[J]. *Nat Mach Intell*, 2021, 3(1):68.
- [17] Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing:progress, challenges and recommendations [J]. *Nat Rev Drug Discov*, 2019, 18(1): 41-58.
- [18] Thomas KL. Predicting new molecular targets for known drugs[J]. *Nature*, 2009, 462(7270):175-181.
- [19] Feinberg EN, Sur D, Wu Z, et al. PotentialNet for molecular property prediction[J]. *ACS Cent Sci*, 2018, 4(11):1520-1530.
- [20] Hochuli J, Helbling A, Skaist T, et al. Visualizing convolutional neural network protein-ligand scoring[J]. *J Mol Graph Model*, 2018, 84(9):96-108.
- [21] Issa NT, Peters OJ, Byers SW, et al. RepurposeVS:a drug repurposing-focused computational method for accurate drug-target signature predictions [J]. *Comb Chem High Throughput Screen*, 2015, 18(8):784-794.
- [22] Wigley FM, Flavahan NA. Raynaud's Phenomenon [J]. *N Engl J Med*, 2016, 375(6):556-565.
- [23] Khouri C, Lepelley M, Bailly S, et al. Comparative efficacy and safety of treatments for secondary Raynaud's phenomenon: a systematic review and network meta-analysis of randomised trials[J]. *Lancet Rheumatol*, 2019, 1(4):E237-E246.
- [24] Zaza P, Matthieu R, Jean-Luc C, et al. Drug repurposing in Raynaud's phenomenon through adverse event signature matching in the World Health Organization pharmacovigilance database[J]. *Br J Clin Pharmacol*, 2020, 86(11):2217-2222.
- [25] Pinzi L, Rastelli G. Molecular docking:shifting paradigms in drug discovery [J]. *Int J Mol Sci*, 2019, 20(18):4331.
- [26] Xu X, Huang M, Zou X. Docking-based inverse virtual screening:methods, applications, and challenges[J]. *Biophys Rep*, 2018, 4(1):1-16.
- [27] Dakshanamurthy S, Issa NT, Assefnia S, et al. Predicting new indications for approved drugs using a proteochemometric method [J]. *J Med Chem*, 2012, 55(15):6832-6848.
- [28] Zahler S, Tietze S, Totzke F, et al. Inverse in silico screening for identification of kinase inhibitor targets[J]. *Chem Biol*, 2007, 14(11):1207-1214.
- [29] Parvathaneni V, Kulkarni NS, Muth A, et al. Drug repurposing:a promising tool to accelerate the drug discovery process[J]. *Drug Discov Today*, 2019, 24(10):2076-2085.
- [30] Smith SB, Dampier W, Tozeren A, et al. Identification of common biological pathways and drug targets across multiple respiratory viruses based on human host gene expression analysis[J]. *PLoS One*, 2012, 7(3):e33174.
- [31] Mejia-Pedroza RA, Espinal-Enriquez J, Hernandez-Lemus E. Pathway-based drug repositioning for breast cancer molecular subtypes[J]. *Front Pharmacol*, 2018, 9:905-917.
- [32] Rivero-Garcia I, Castresana-Aguirre M, Guglielmo L, et al. Drug repurposing improves disease targeting 11-fold and can be augmented by network module targeting, applied to COVID-19 [J]. *Sci Rep*, 2021, 11(1):20687.
- [33] Sansau P, Agarwal P, Barnes MR, et al. Use of genome-wide association studies for drug repositioning[J]. *Nat Biotechnol*, 2012, 30(4):317-320.
- [34] Nabirotkin S, Peluffo AE, Rinaudo P, et al. Next-generation drug repurposing using human genetics and network biology[J]. *Curr Opin Pharmacol*, 2020, 51:78-92.
- [35] Fang J, Zhang P, Wang Q, et al. Artificial intelligence framework identifies candidate targets for drug repurposing in Alzheimer's disease[J]. *Alzheimer's Research & Therapy*, 2022, 14(1):7.
- [36] Wang Q, Xu R. Drug repositioning for prostate cancer:using a data-driven approach to gain new insights[J]. *AMIA Annu Symp Proc*, 2017, 2017: 1724-1733.
- [37] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence - What is it and what can it tell us? [J]. *N Engl J Med*, 2016, 375(23): 2293-2297.
- [38] Hayrinen K, Saranto K, Nykanen P. Definition, structure, content, use and impacts of electronic health records;a review of the research literature [J]. *Int J Med Inform*, 2008, 77(5):291-304.
- [39] Xu H, Aldrich MC, Chen Q, et al. Validating drug repurposing signals using electronic health records;a case study of metformin associated with

- duced cancer mortality [J]. J Am Med Assoc, 2015, 22(1): 179-191.
- [40] Ribelles N, Jerez JM, Rodriguez-Brazzarola P, et al. Machine learning and natural language processing (NLP) approach to predict early progression to first-line treatment in real-world hormone receptor-positive (HR+)/HER2-negative advanced breast cancer patients [J]. Eur J Cancer, 2021, 144: 224-231.
- [41] Frost J, Okun S, Vaughan T, et al. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe [J]. J Med Internet Res, 2011, 13(1): e6.
- [42] Liu Y, Kohlberger T, Norouzi M, et al. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists [J]. Arch Pathol Lab Med, 2019, 143(7): 859-868.
- [43] Wang F, Kaushal R, Khullar D. Should health care demand interpretable artificial intelligence or accept "black box" medicine? [J]. Ann Intern Med, 2020, 172(1): 59-60.
- [44] Lee CY, Chen YP. New insights into drug repurposing for COVID-19 using deep learning [J]. IEEE Trans Neural Netw Learn Syst, 2021, 32(11): 4770-4780.
- [45] Shingjergji K, Celebi R, Scholtes J, et al. Relation extraction from DailyMed structured product labels by optimally combining crowd, experts and machines [J]. J Biomed Inform, 2021, 122: 103902.
- [46] Azencott CA. Machine learning and genomics: precision medicine versus patient privacy [J]. Philos Trans A Math Phys Eng Sci, 2018, 376(2128): 20170350.
- [47] 王旭. 大数据医疗时代的人工智能与隐私保护 [J]. 电子产品世界, 2019, 26(6): 79-81.
- [48] Hamet P, Tremblay J. Artificial intelligence in medicine [J]. Metabolism, 2017, 69S: S36-S40.
- [49] Hessler G, Baringhaus KH. Artificial intelligence in drug design [J]. Molecules, 2018, 23(10): 2520.

(收稿日期: 2022-03-30)

《新型冠状病毒(2019-nCoV)核酸检测试剂注册审查指导原则》适用范围

为进一步规范新型冠状病毒相关检测试剂的管理,国家药品监督管理局器审中心组织制定并于2022年4月27日发布了《新型冠状病毒(2019-nCoV)核酸检测试剂注册审查指导原则》(以下简称《指导原则》)。

适用范围

新型冠状病毒(以下简称“新冠病毒”)属于 β 属冠状病毒,有包膜,颗粒呈圆形或椭圆形,直径约为60~140 nm。具有5个必需基因,分别针对核蛋白(N)、病毒包膜(E)、基质蛋白(M)和刺突蛋白(S)4种结构蛋白及RNA依赖性的RNA聚合酶(RdRp)。核蛋白(N)包裹RNA基因组成核衣壳,外面围绕着病毒包膜(E),病毒包膜包埋有基质蛋白(M)和刺突蛋白(S)等蛋白。

实验室检查包括一般检查,病原学及血清学检查,胸部影像学检查等。病原学检查又包括核酸检测和抗原检测。核酸检测主要采用逆转录PCR,二代测序等方法,在鼻、口咽拭子、痰液和其他下呼吸道分泌物等标本中均可检测出新冠病毒核酸。

新冠病毒在流行过程中基因组不断发生变异,新的变异株可能在传播力、致病性、免疫逃逸能力等方面发生改变。变异株可能影响检测试剂的性能,甚至出现漏检。

本《指导原则》适用于采用逆转录实时荧光PCR法,对咽拭子、鼻咽拭子、肺泡灌洗液或痰液等呼吸道样本中的新冠病毒核酸进行体外定性的检测试剂。

对于采用其他方法学的新冠病毒核酸检测试剂,可能部分要求不完全适用或本文所述内容不够全面,申请人应参照本指导原则,根据产品特性对适用部分进行评价,并补充其他的评价资料。

本《指导原则》适用于新冠病毒核酸检测试剂注册申请和变更注册申请的情形。本《指导原则》针对新冠病毒核酸检测试剂注册申报资料中的部分内容进行撰写,其他未尽事宜应当符合《关于公布体外诊断试剂注册申报资料要求和批准证明文件格式的公告》(国家药品监督管理局公告2021年第122号)等相关法规要求。

(来源:国家药品监督管理局器审中心网站)