molecules

MDPI

审查

用于药物靶点相互作用预测的机器学习

Ruolan Chen [1], Xiangrong Liu [1], Shuting Jin [1], Jiawei Lin [1] and Juan Liu [2,*]

1. 厦门大学信息科学与技术学院计算机科学系，中国厦门 361005；chenruolan@stu.xmu.edu.cn（陈若兰）；xrliu@xmu.edu.cn（刘晓蓉）；stjin.xmu@gmail.com（金思婷）；23020161153321@stu.xmu.edu.cn（林佳琳）
2. 厦门大学航空航天学院仪器与电气工程系，中国厦门 361005

***** 通信作者：cecylui@xmu.edu.cn

收稿日期：2018 年 8 月 5 日；接受日期：2018 年 8 月 27 日；发表日期：2018 年 8 月 31 日

摘要：识别药物与靶点的相互作用将极大地缩小候选药物的筛选范围，因此可作为药物发现过程中的关键第一步。鉴于体外实验成本高昂且耗时，高效计算预测方法可作为药物靶点相互作用（DTI）预测的有前景策略。在本综述中，我们的目标是聚焦于机器学习方法，并提供全面概述。首先，我们总结了药物发现中常用的数据库简要列表。接下来，我们采用分层分类方案，介绍每个类别中的几种代表性方法，尤其是近期最先进的方法。此外，我们比较了每个类别中方法的优势和局限性。最后，我们讨论了机器学习在 DTI 预测中面临的挑战和未来展望。本文可为未来研究者提供基于机器学习的 DTI 预测的参考和教程见解。

关键词：药物靶点相互作用预测；机器学习；药物发现

1. 简介

大多数药物通过与体内靶分子（如酶、离子通道、核受体和 G 蛋白偶联受体（GPCRs））的相互作用来发挥药效。因此，识别药物 - 靶点相互作用（DTIs）已成为包括多药理学、药物再定位、药物发现、副作用预测和耐药性等领域的关键前提条件[1]。药物 - 靶点对的实验和确认一直是许多药物研究的巨大障碍。此外，对于尚未发现的药物 - 靶点相互作用进行生化实验，不仅成本高昂，而且耗时且具有挑战性。例如，每种新分子实体（NME）的研发成本约为 18 亿美元[2]，而新药申请（NDA）的平均审批时间则为 9 至 12 年[3]。

除了已知的存储在各种数据库中的相互作用之外，还有无数未配对的小分子化合物有可能被发现并开发成新的药物。在当前的数据集中，只有少量的药物-靶点对经过了实验验证。实际上，尽管 PubChem 数据库中描述的化合物超过 9000 万种，但仍有很大比例的相互作用有待发现[4]。此外，尽管生物技术取得了进步，但近年来获得监管机构批准的真正创新药物数量却在减少。例如，据报道，美国食品药品监督管理局（FDA）每年仅批准约 20 种新药，且研发成本高昂[5]。这些巨大的时间、资金和人力物力成本促使研究人员不断开发新的方法。

用于开发新药的创新技术。相互作用预测有助于高效筛选新药候选药物。

为现有或已弃用的药物确定新的作用靶点，即药物再定位，是药物研发中的另一个重要环节。随着我们对药理学理解的不断深入，“多靶点、多药物”模式已取代“一靶点、一药物”模式而被广泛接受[1]。一个重要的事实是，药物通常作用于多个蛋白质，而非仅仅一个。抗癌药物舒尼替尼（Sutent）和伊马替尼（Gleevec）就是有力的例证。此外，药物除了主要的治疗靶点外，还可能与其他蛋白质相互作用，即脱靶效应。脱靶效应通常被视为有害的副作用。然而，在某些情况下，它们可能是有益的，因为它们可能会带来意想不到的治疗效果，并为药物副作用的分子机制提供新的视角。药物再定位的目的是发现现有药物的新临床用途。药物再定位的一个明显优势在于，现有药物的安全性和生物利用度已得到严格验证。省去一些先前已完成的步骤可以大大加快药物研发进程。近来，世界各地的政府、学术机构和非营利组织在药物再定位方面投入了更多精力，这将极大地促进药物再定位的研究[6]。

鉴于上述所有原因，检测药物与靶点的相互作用对于新药研发和老药新用都至关重要。基于湿实验的已知药物与靶点相互作用的数量非常有限。已知与未知药物靶点对之间的巨大差距促使人们关注药物靶点相互作用（DTI）的预测。传统的体外预测策略面临着时间和资金成本的限制，而最近开发的计算或计算机模拟方法能够更高效地预测潜在的相互作用候选者。计算方法在许多相关生物信息学领域都取得了良好的效果，例如疾病相关 miRNA 预测[7-9]、疾病基因预测[10]、蛋白质-蛋白质相互作用预测[11]和蛋白质亚细胞定位预测[12]。它们极大地缩小了实验验证药物靶点相互作用的研究范围。因此，对药物靶点相互作用预测的计算技术的开发有着持续且迫切的需求。

目前，基于配体、对接模拟和化学基因组学的方法是预测药物靶点相互作用（DTIs）的三大主要计算方法。基于配体的方法，如定量构效关系（QSAR），利用相似分子通常与相似蛋白质结合这一理念。具体而言，这些方法通过将新的配体与已知的蛋白质配体进行比较来预测相互作用。然而，当已知配体的数量不足时，基于配体的方法表现不佳。

至于对接模拟方法[14]，由于其需要蛋白质的三维（3D）结构来进行模拟，所以在大量蛋白质的 3D 结构无法获取的情况下就无法适用。此外，对于像离子通道和 G 蛋白偶联受体（GPCRs）这类结构过于复杂的膜蛋白，这种方法也无法应用。对接模拟通常需要耗费大量时间，因此效率尤其低下。

为解决传统方法的难题，化学基因组学方法[15]近来在大规模药物发现和再定位中已成功应用。在药物靶点相互作用（DTI）预测中，通常涉及四种主要类型的靶点，即蛋白质、疾病、基因和副作用。为了预测药物-靶点对，这些方法将化合物的化学空间和靶蛋白的基因组空间整合到一个统一的空间：药理空间。因此，化学基因组学方法能够充分利用有利于预测的大量生物数据。在这样的 DTI 预测问题中，主要挑战在于已知药物-蛋白质相互作用的稀缺性以及未经验证的负药物-靶点相互作用样本的缺乏。这些化学基因组学方法可归为不同类别，如基于机器学习的方法、基于图的方法和基于网络的方法[16]。在所有化学基因组学方法中，基于机器学习的方法因其可靠的预测结果而备受关注。这些方法大多利用药物和靶点的化学及生物学特征，并采用

多种机器学习技术用于预测药物与靶点之间的相互作用。图 1 是近期用于药物 - 靶点相互作用预测的计算方法分支图。

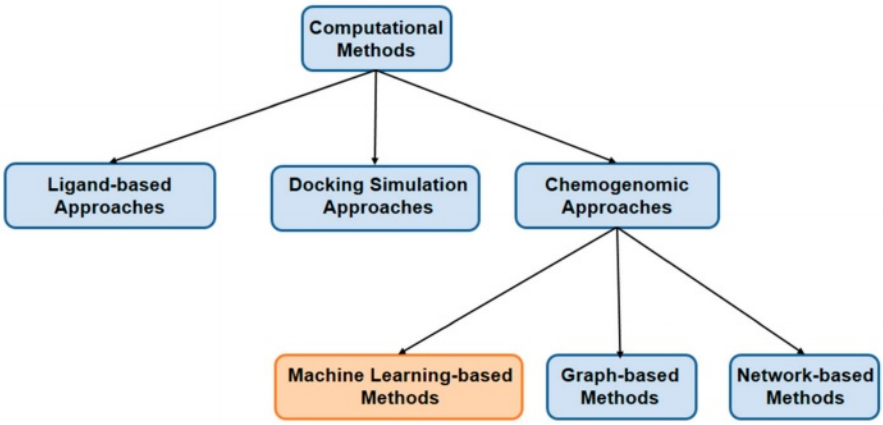

图 1.近期用于弥散张量成像（DTI）预测的计算方法分支图。

在这篇综述中，我们重点关注应用于药物靶点相互作用（DTI）预测的机器学习方法。具体而言，我们旨在全面概述利用机器学习框架的化学基因组学方法的一个子类。与那些同样应用机器学习策略的基于配体的方法相比，本综述中讨论的方法适用于已知配体不足的靶蛋白。首先，我们简要总结了药物发现中常用的数据库列表。接下来，我们采用分层分类方案。特别是，我们将机器学习方法分为两大类，即监督学习和半监督学习方法，并提供更多的子类。我们试图分别介绍每类中的几种代表性方法。此外，我们还将阐述每类方法的优势和劣势。最后，我们将从我们的角度讨论当前机器学习方法在 DTI 预测领域面临的挑战和未来展望。

1. 监督学习方法 在训练集中需要同时有正标签和负标签。然后利用这些带标签的样本训练学习模型，以用于后续的药物 - 靶点相互作用预测。

   · 基于相似性的方法 通过各种相似性度量策略计算药物之间或靶点之间的相似性。相似性矩阵可用于多种类型的核函数：

     （i）最近邻方法：最近邻方法是基于最近邻的信息来进行预测的。

     （ii）二分局部模型：首先分别针对药物和靶点训练两个局部模型。每个药物 - 靶点对的最终预测结果是基于这两个独立预测得分的运算得出的。

     （iii）矩阵分解方法：将药物 - 靶点相互作用矩阵分解为两个潜在特征矩阵，这两个矩阵相乘可近似得到原始矩阵。

   · 基于特征向量的方法 将训练数据表示为特征向量。然后可以利用一些机器学习模型，如随机森林，基于这些向量进行预测。

2. 半监督学习方法 半监督学习方法仅基于少量有标签数据和大量无标签数据进行预测。据我们所知

在知识方面，已经有一些关于化学基因组学方法预测DTI的优秀综述[6,15 - 19]。与以前的工作相比，我们专注于DTI预测中使用的机器学习方法的特殊主题。此外，我们采用了分层分类方案，并总结了几种最新的预测方法，如[20-23]，这些方法在以往的综述中很少被提及。特别地，评论[17]只是从一个狭窄的角度写的，即基于相似性的方法，这是机器学习方法的一个子类。调查[6,15,18,19]都提供了对化学基因组学方法的更一般和全面的概述，而不是强调机器学习。近年来，机器学习取得了突破性进展，引起了公众的广泛关注。从这个特殊的角度讨论最新的DTI预测策略可以展示更多的方法细节。虽然评论[16]也关注基于学习的方法，但它的重点只是监督学习。相比之下，我们在回顾[16]发表后提供了更详细的子类并介绍了新开发的方法。本文其余部分的结构安排如下：“数据库”部分介绍了当前可用于药物 - 靶点相互作用预测研究的数据源。“方法”部分通过分层分类方案简要介绍了几种具有代表性的机器学习方法。然后，我们讨论了每一类方法的优势和局限性以及仍存在的挑战。最后，“结论与展望”部分对机器学习在药物 - 靶点相互作用预测中的未来前景进行了展望。

2. 数据库

基于现有生物信息学数据库的数据挖掘和利用是药物发现的一种重要方法。随着分子生物学的发展，有关药物和靶点的大量信息不断积累。因此，有必要建立数据库来管理和维护这些数据。到目前为止，已经存在许多涉及各种化学化合物家族潜在细胞靶点的不同专业数据库。其中很大一部分是公开可用的。此外，由于世界各地研究人员的贡献，数据量不断增加。随着有关药物和靶点的信息收集越来越多，药物发现研究的机会也越来越多。在一定程度上，这些数据库促进了最新药物发现方法的发展。在表 1 中，我们列出了常用的数据库、其网络服务器和简要描述。表 2 显示了这些数据库中化合物、靶点和化合物 - 靶点相互作用的数量统计。请注意，并非所有数据库在其数据库和已发表的论文中都提供了完整的信息。

其中一些数据库经常更新，例如 DrugBank、KEGG 和 STITCH 等，而其他一些数据库的数据多年几乎未变，比如 SuperPred 最后一次更新是在 2014 年 4 月。不过令人欣慰的是，最近建立了更多新的数据库和易于使用的网络服务器。一方面，现有的数据库提供了丰富的药物空间和靶点空间的数据来源。现在是研究人员努力整合更多不同类型异构数据的时候了。另一方面，当前的数据库都不涉及任何药物与靶点之间无相互作用的信息。这一共同缺陷限制了监督学习方法的预测结果。因此，未来公开药物与靶点之间的相互作用和无相互作用信息将具有重要意义。

表 1.支持药物发现方法的数据库。

数据库和网址	简要描述
KEGG [29] http://www.genome.jp/kegg	一个基因和基因组的百科全书，既用于基因组信息的功能解读，也用于其实际应用。
布伦达 [30] http://www.brenda-enzymes.org/	主要的酶和酶-配体信息库。
PubChem [31] https://pubchem.ncbi.nlm.nih.gov/ （注：此网址无需翻译）	一个关于化学物质及其生物活性信息的数据库，包含三个相互关联的数据库，即物质、化合物和生物测定数据库。
TTD [32] http://bidd.nus.edu.sg/group/ttd/ttd.asp	治疗靶点数据库提供有关耐药性突变、基因表达和靶点组合数据的全面信息。
DrugBank [33] http://www.drugbank.ca	该数据库由两部分信息组成，分别涉及详细的药物数据（即化学、药理和制药方面）以及药物靶点信息（即序列、结构和通路）。
SuperTarget [34] http://bioinf-apache.charite.de/supertarget	一个整合了药物相关信息的数据库，包含超过 33 万种化合物与靶点蛋白的关系。
ChEMBL [35] https://www.ebi.ac.uk/chembldb	定期从已发表的原始文献中收集的分子结构及分子 - 蛋白质相互作用的数据资源。
STITCH [36] http://stitch.embl.de/ （注：此网址无需翻译，直接保留原样。）	已知和预测的化学物质 - 蛋白质相互作用库。
MATADOR [37] http://matador.embl.de/ （注：此网址为英文网站，无需翻译。）	一个包含尽可能多的直接和间接蛋白质 - 化学物质相互作用的数据库。
BindingDB [38] http://www.bindingdb.org/bind	一个蛋白质 - 配体结合亲和力的公共数据库。
TDR 目标 [39] http://tdrtargets.org/	一种针对被忽视的热带病的化学基因组学资源。
SIDER [40] http://sideeffects.embl.de/ （注：此网址为英文网站，无需翻译。）	提供上市药品的信息及其记录的药品不良反应情况。
ChemBank [41] http://chembank.broad.harvard.edu/ （注：此网站链接无需翻译，直接保留原样即可。）	来自小分子及小分子筛选的可用数据集以及用于研究其特性的资源。
DCDB [42] http://www.cls.zju.edu.cn/dcdb/ （注：此网址为英文原文中的链接，未做翻译。）	药物组合数据库，用于收集和整理已知的药物组合实例。
CancerDR [43] http://crdd.osdd.net/raghava/cancerdr/ （注：此链接指向的是一个网站，通常在翻译时会保留原链接，因此此处未做翻译处理。）	包含 148 种抗癌药物及其对约 1000 种癌细胞系疗效的癌症药物耐药性数据库。
ASDCD [44] http://asdcd.amss.ac.cn/ （由于该网址为链接，且未提供具体需要翻译的内容，所以仅保留原文。）	首个抗真菌协同用药组合数据库，涵盖已发表的抗真菌药物协同组合、作用靶点、适应症及其他相关数据。
SuperPred [45] http://prediction.charite.de/ （注：此网址无需翻译，直接保留原样。）	化合物-靶点相互作用的资源。

表 2.综述中所涵盖数据库中化合物、靶点及化合物 - 靶点相互作用的数量统计。

数据库	化合物的数量	目标数量	化合物与靶点的相互作用数量
KEGG	18,380	26,885,475	
布伦达		7341	
PubChem	96,479,316	68,868	
TTD	34,019	3101	
DrugBank	11,682	26,889	131,724
超级目标	195,770	6219	332,828
ChEMBL	2,275,906	12,091	
针	500,000	9,600,000	1,600,000,000
斗牛士	775		
BindingDB	652,068	7082	1,454,892
TDR 目标	2,000,000	5300	
SIDER （无准确对应的中文词汇）	5868	1430	139,756
ChemBank	1,700,000		
DCDB	904	805	
癌症DR	148	116	
ASDCD	105	1225	210
超级预测器	341,000	1800	665,000

3. 方法

在大数据时代，机器学习方法旨在基于某种底层算法和给定的大数据集生成预测模型。对于生物和生物医学研究而言，机器学习在将大量数据筛选为模式方面发挥着关键作用[24-27]。药物-靶点相互作用（DTI）预测中的一般机器学习工作流程可分为三个步骤。首先，对药物和靶点的输入数据进行预处理；其次，基于一组学习规则训练底层模型；最后，利用预测模型对测试数据集进行预测。

通过我们的研究发现，文献[28]是首次将机器学习应用于蛋白质与化学物质相互作用预测的工作。该研究建立了基于氨基酸序列数据、化学结构数据和质谱数据的 SVM 分析框架。这项开创性的研究启发了后续的研究。自此，机器学习在药物发现中的应用已成为一个长期且日益受到关注的领域。

为简便起见，我们将用于药物 - 靶点相互作用预测的机器学习方法分为两大类，即监督学习和半监督方法。具体而言，监督学习方法又可进一步细分为基于相似性的方法和基于特征的方法两类。

3.1. 监督学习方法

在有标签的情况下，监督学习方法用于训练学习模型并识别模式。对于药物-靶点相互作用（DIT）预测问题，已知的药物-靶点相互作用被标记为正样本，其余的则被标记为负样本。接下来，这些标签用于训练模型以进行后续的相互作用预测。实际上，那些没有明确相互作用信息的药物-靶点对可能对应于未知或缺失的相互作用，而非

药物与靶点之间未发生相互作用的情况。通常，药物与靶点之间未发生相互作用的结果不会被发表。这类方法将所有未知的药物 - 靶点相互作用都视为未发生相互作用，尽管这种做法并不准确。在本节中，我们将回顾迄今为止提出的两类有监督方法，即基于相似性的方法和基于特征的方法。

3.1.1. 基于相似性的方法

基于相似性的机器学习方法的一个关键潜在假设是“连坐”假设，即相似的药物往往具有相似的作用靶点，反之亦然。在这种方法中，药物之间或靶点之间的相似性通过各种相似性度量来计算。所构建的相似性矩阵定义了几种类型的核函数。

・最近邻方法

最近邻方法通常采用相对简单的相似性函数。研究人员经常将这些方法与其他一些方法相结合，以帮助预测新药或新靶点，例如文献[46,47]中的模型。在早期阶段，研究[48]提出了两种探索性方法，即最近邻谱方法（NN）和加权谱方法。最近邻谱方法遵循这样一个关键概念，即相似的药物或靶点在网络中往往距离较近。该方法在[49]中被用作基准。相比之下，加权谱方法利用所有其他药物和靶点的相似性，然后采用加权平均。然而，当结合相似药物的靶点具有较低的序列相似性或反之亦然时，这些方法表现不佳。

在张等人开展的研究[23,50]中，开发了基于邻域的药物-药物配对预测方法。这些研究进一步将经典的邻域推荐方法扩展为基于集成邻域的方法（INBM）。简单来说，邻域推荐方法通常使用邻域的加权平均信息进行预测。INBM 是一种集成模型，它整合了多个基于邻域的模型以实现稳健的预测。对于每一对药物，使用三种常用的公式，即杰卡德相似度、余弦相似度和皮尔逊相关相似度，来计算相似度得分。

此类方法中的另一种新颖方法是基于相似性排序的预测器（SRP）[51]。通过计算两个指标，即倾向指数和逆倾向指数，来构建 SRP。具体而言，前者表示每个药物 - 靶点对相互作用的可能性，而后者衡量每个药物 - 靶点对不相互作用的倾向。计算公式涉及相似性和相似性排序。然后计算交互可能性得分，作为这两个指标的似然比。该方法可以从药物侧和靶点侧分别生成两个交互可能性得分，最终预测得分是这两个得分的平均值。SRP 的明显优势在于它是一种懒惰且非参数模型，无需优化求解器、先验统计知识以及可调参数。

近年来，其他基于相似性的新方法相继被提出，例如基于规则的推理。由于先前基于拓扑的方法存在局限性，一种基于相似性的深度学习方法[52]将相似性度量与两种基于规则的推理方法相结合。换句话说，采用基于药物的相似性推理（DBSI）和基于靶点的相似性推理（TBSI）[48,53]来发现具有相似性的药物-靶点相互作用。尽管可以灵活组合任何核函数，但该方法无法预测新的药物或靶点。

请注意，大多数相似性度量仅利用一些重要的药物相关或疾病相关属性来进行药物 - 疾病预测，并且忽略了已知的药物 - 疾病相互作用信息 [54]。一些研究人员提出了新的相似性度量方法。罗等人 [54] 设计了一种综合相似性度量方法。为了改进用于药物 - 疾病预测的传统相似性度量，该综合相似性度量将药物或疾病特征信息与已知的药物 - 疾病相互作用相结合。该相似性度量可以分为三个步骤。在第一步中，分别基于药物相关属性或疾病相关属性计算药物相似性和疾病相似性。在第二步中，这些相似性值被

根据分析和评估结果，通过逻辑函数进行调整。在最后一步，可以为药物相似性建立加权药物网络。边权重表示相应药物之间共同疾病的数量。然后应用聚类方法 ClusterONE 来识别潜在的药物簇。属于同一簇的药物之间的相似性得到增强，从而获得全面的药物相似性。疾病相似性也可以以与药物相同的方式得到改进。

· 二分局部模型

二部图局部模型（BLMs）首先分别独立生成药物和靶点的两个预测结果。最终的预测结果则是通过将这两个预测得分进行聚合而获得。

BLM 这一概念最早由 Bleakley 和 Yamanishi 在其开创性研究中提出[49]。该方法能够将药物 - 靶点相互作用预测问题转化为二分类问题。更具体地说，基于化学相似性为药物训练一个局部模型，基于序列结构为蛋白质训练另一个局部模型。因此，两个支持向量机（SVM）分类器能够分别从药物或靶点的角度生成两个独立的预测结果。对于每一对药物 - 靶点，最终的预测结果是基于这两个独立预测得分的平均值计算得出的。

类似地，另一种方法[55]开发了一种正则化最小二乘分类器，引入了两种算法，分别称为 RLS-avg 和 RLS-kron。特别是，正则化最小二乘法（RLS-avg）利用核岭回归进行预测。而在 RLS-kron 中，将所有药物和靶点的组合视为一个，进行克罗内克积运算，从而大大降低了运行时间。

鉴于上述基于二分图局部模型（BLM）的方法在没有已知相互作用的情况下难以预测新药物或新靶点的局限性，Mei 等人[46]通过添加一个预处理步骤来扩展现有的 BLM，该步骤从邻居的相互作用谱中推断训练数据。该方法被称为基于邻居相互作用谱推断的二分图局部模型（BLM-NII）。BLM-NII 涉及 RLS-avg 算法，并被证明在新候选问题上是有效的。

· 矩阵分解方法

矩阵分解方法通常用于推荐系统中，以发现潜在的用户 - 项目交互。药物 - 靶点相互作用预测可以被视为一个矩阵补全问题，旨在寻找缺失的相互作用。因此，药物 - 靶点相互作用矩阵可以分解为另外两个矩阵，这两个矩阵相乘可以近似于原始矩阵。

具有孪生核的核化贝叶斯矩阵分解（KBMF2K）[56] 是将矩阵分解引入药物-靶点相互作用预测的原始方法。与一些先前的方法类似，KBMF2K 仅基于药物化合物之间的化学相似性和靶点蛋白质之间的基因组相似性来定义两个核矩阵。它结合了贝叶斯概率公式、矩阵分解和二分类方法来解决预测问题。

另一项采用概率公式的研究是概率矩阵分解（PMF）[57]。PMF 与 KBMF2K 的显著区别在于其不依赖药物或靶点相似性矩阵。此外，该研究还提出了与概率矩阵分解相结合的主动学习（AL）策略。

郑等人[58]提出了一种从一类协同过滤（CMF）扩展而来的加权低秩近似方法，即多相似性协同矩阵分解（MSCMF）。MSCMF 集成了多种相似性矩阵，包括化学结构相似性、基因组序列相似性、ATC 相似性、GO 相似性和 PPI 网络相似性。通过交替最小二乘算法估计矩阵权重，从而自动选择相似性。在实验中，这种策略提高了预测性能。药物和靶点被投影到低秩矩阵中。然而，尽管其性能良好，但在这种数据集成策略下，可能会丢失大量信息，从而导致次优解。

Ezzat 等人[59]开发的方法采用了两种矩阵分解方法（即 GRMF 和 WGRMF）。先前的研究[60]表明，数据通常位于或接近低维非线性流形。因此，GRMF 和 WGRMF 通过图正则化隐式地执行流形学习。此外，在新药物或靶点预测中应用了一个预处理步骤（WKNKN），即将原始药物-靶点矩阵中的所有 0 转换为相互作用可能性值。这一重要步骤使该方法有别于其他将给定药物-靶点矩阵中的所有 0 粗略视为无相互作用的方法，从而提高了预测结果。

3.1.2. 基于特征向量的方法

通常，基于相似性的预测算法并未考虑语义网络中定义的异构类型和相互作用。此外，添加两个节点之间的长间接连接可能会很困难。因此，基于特征向量的方法已被用于药物 - 靶点相互作用（DTI）预测。基于特征向量的方法的输入是通过固定长度的特征向量表示的药物 - 靶点对。这些特征向量由药物和靶点的各种属性进行编码。

在系统方法中[61]，使用 DRAGON 程序（http://www.talete.mi.it/index.htm）计算化学描述符。最终，每种药物都由包括组成描述符、拓扑描述符、二维自相关、基于特征值的指标等在内的 1080 个描述符来表示。同样，通过 PROFEAT WEBSEVER（http://jing.cz3.nus.edu.sg/cgi-bin/prof/prof.cgi）将每个蛋白质表示为一组结构和物理化学描述符，这些描述符包括氨基酸组成描述符、二肽组成描述符和自相关描述符等。这样，每个长度可变的蛋白质序列都可以转换为 1080 维的标准特征向量。因此，可以为每个药物 - 靶点对构建一组 2160 维的特征向量。后续的预测步骤执行随机森林（RF）算法，该算法在树中引入随机训练集（自助法）和随机输入向量。在实验中，这种综合框架显示出其对过拟合问题的稳健性，并且对于大规模数据集的处理效率更高。

为了整合来自异构数据源的多样化信息，罗等人[20]提出了一种名为 DTINet 的方法。通过 DTINet，首先学习到一个低维特征向量，该向量能准确解释异构网络中每个节点的拓扑属性。在后续步骤中，DTINet 应用归纳矩阵补全，以最佳方式将药物空间投影到蛋白质空间。

由于 DTINet 会分离特征，可能会导致最优解的丢失，Wan 等人[21]创建了一个新的框架，称为用于药物-靶点相互作用预测的邻域信息神经整合（NeoDTI）。NeoDTI 的灵感来源于卷积神经网络（CNNs）。它整合了异质网络中的邻域信息。在提取药物和靶点的复杂隐藏特征向量之后，NeoDTI 自动学习拓扑保持表示，从而实现更优的预测性能。

文献[62]率先将一种两层无向图模型，即受限玻尔兹曼机（RBM）引入大规模药物-靶点相互作用预测。这些层内不存在连接。此外，RBM 模型通过一种实用的学习算法，即对比散度（CD）进行训练。该方法显著优于其他现有方法之处在于，它能够在多维网络上预测不同类型的药物-靶点相互作用。换句话说，该方法不仅能识别二元药物-靶点相互作用，还能确定其相应的相互作用类型，包括关系和药物作用模式。

在傅及其合作者发表的论文[63]中，基于元路径的拓扑特征构建了一个先进的机器学习模型。计算了两种拓扑特征度量，包括节点间路径实例的数量以及对其进行的归一化处理。给定这些特征后，采用随机森林算法进行监督分类。此外，还探讨了其内在的

随机森林中嵌入的特征排序算法会选择重要的拓扑特征以实现更精准的预测。该框架已展现出精确的预测能力。

3.2. 半监督学习方法

鉴于负样本的选择对 DTI 预测结果的准确性有很大影响，一些研究人员提出了半监督方法来解决这一问题。这些方法仅使用少量的有标签数据和大量的无标签数据。半监督方法通常利用有标签数据来推断无标签数据的标签。另一方面，无标签数据也有助于提供有关训练集结构的见解。

由于没有负样本可用，研究 [64] 首次采用了基于 BLM 概念的流形拉普拉斯正则化最小二乘法（LapRLS）。此外，还提出了标准 LapRLS 的一种扩展方法，即 NetLapRLS。NetLapRLS 将化学空间、基因组空间和药物 - 蛋白质相互作用的信息整合到一个新的核函数中。这些半监督方法所取得的结果比仅使用有标签数据时更令人鼓舞。然而，在大规模实施时，这些方法会耗费大量时间。

另一种方法适用于半监督和无监督场景。马等人[22]提出了一种新框架，用于在标签稀缺的情况下学习准确且可解释的相似性度量。该框架构建了一组基于图自编码器（GAE）的模型，并整合了多视图药物相似性。此外，还使用了注意力机制来进行视图选择，以提高可解释性。

3.3. 讨论

每种机器学习模型都有其独特的优势和劣势。请注意，正如计算机科学中广为人知的“没有免费的午餐定理”[65] 所述，机器学习方法具有特定的适用情境。因此，在本综述中，我们只能基于药物靶点相互作用预测的情境来评估每种方法类别的优缺点。

已有多项有监督模型被证实可用于药物 - 靶点相互作用（DTI）预测。然而，大多数有监督方法只是将所有未标记的药物 - 靶点对视为负样本，从而导致预测结果不准确。此外，由于相似性矩阵计算的复杂度较高，每种基于相似性的方法在扩展到大型数据集时都存在局限性。

分别考虑基于相似性的三种子类方法。尽管最近邻方法通常采用相对简单的相似性函数，但它们中的大多数仅基于一阶相似性构建邻域，而不涉及相似性的传递性[66]。二分图局部模型的一个关键优势在于，它们处理的药物 - 靶点对要少得多，因此其复杂度远低于全局模型。然而，除非与其他方法结合，否则二分图局部模型无法处理药物和靶点均未出现在训练集中的情况。根据[19]中的实验结果，矩阵分解方法通常比包括最近邻模型和二分图局部模型在内的其他方法具有更优越的性能。

已知的药物靶点相互作用数量较少，导致数据集不平衡。作为处理不平衡数据集的有效方法，半监督学习仅使用少量有标签数据和大量无标签数据，就能生成比监督学习更可靠的预测结果。

除了上述提到的单机学习方法，我们还引入了几种集成方法[61,63]。通常，通过不同单机方法偏差的相互抵消，能够得到更优且更稳健的预测结果。一般来说，集成方法可以将不同的学习模型结合起来。关于更多应用于药物-靶点相互作用预测任务的集成方法，请参阅[67-69]。

总体而言，机器学习在药物靶点相互作用（DTI）预测方面取得了良好的效果。然而，仍存在诸多挑战。首先，近来一些研究人员强调指出

基于机器学习的预测模型通常是在过于简化的设置下建立和评估的。在这样的实验设置下得出的预测结果可能过于乐观，与实际情况存在偏差。特别是，大多数机器学习方法简单地将药物 - 靶点相互作用视为开 - 关关系，而忽略了分子浓度和定量亲和力等其他重要因素。帕希卡拉等人[24]指出了对预测结果有显著影响的四个因素，包括问题的表述、评估数据集、评估程序和实验设置。考虑到药物 - 靶点对的结合亲和力和剂量依赖性，药物 - 靶点相互作用预测问题应表述为回归或排序预测问题，而非标准的二分类问题。第二个挑战是数据集不平衡的问题。由于已知药物 - 靶点对的数量较少，当前的数据集是不平衡的。一些模型，如决策树和支持向量机，对识别多数类存在很大的偏差，从而导致性能不佳[16]。第三，大多数机器学习模型具有“可解释性差”的特点。换句话说，从生物学角度很难理解其潜在的药物作用机制。需要注意的是，在大多数情况下，相对简单的模型更容易解释。这种情况与一条“经验法则”[70]相符，即“简单往往更好”。然而，对于大多数目前最先进的、能实现高药物-靶点相互作用（DTI）预测准确率的方法，比如深度学习方法，从药理学角度对其进行解释却很困难。最后但同样重要的是，目前还没有专门针对 DTI 预测的统一评估指标。以往的研究采用了生物信息学中的一些常见评估指标[71]，如灵敏度、特异性、精确率-召回率曲线下的面积（AUPR）和受试者工作特征曲线下的面积（AUC）。事实上，如果灵敏度提高，特异性就会降低。鉴于单独使用灵敏度或特异性存在局限性，AUPR 和 AUC 可能是评估任务中的更好选择。在目前可获取的数据集中，未知样本的数量远多于已知样本，因此应更重视假阳性结果。AUPR 能尽可能降低假阳性数据对评估结果的影响[72]，而 AUC 对不平衡数据集不敏感[73]。因此，AUPR 和 AUC 通常都是评估基于机器学习的方法性能的合适指标。

4. 结论与展望

药物靶点相互作用（DTIs）有助于潜在药物的选择，从而有效缩小生化实验的研究范围。此外，它们还能深入揭示药物的副作用和作用机制。因此，DTI 预测是药物发现的重要前提。事实上，已建立了多个公开可用的数据库，并推动了创新 DTI 预测策略的发展。

在这篇综述中，我们重点关注基于机器学习的整合化学空间和基因组空间的方法。我们总结了在药物靶点相互作用（DTI）预测中常用的数据库和机器学习方法。特别是，我们着重介绍了近年来出现的几种最先进的预测模型。我们采用了一种分层分类方案。我们将机器学习方法分为两大类：有监督和半监督方法，并提供了更多的子类。

在未来几年，机器学习在药物靶点相互作用预测方面将大有可为。然而，仍有很大的改进空间。因此，我们在此提出一些建议，供未来的研究人员参考。

首先，集成方法将多个独立的分类器组合成一个模型，通常能取得更好的预测结果。其次，半监督学习是解决数据集不平衡问题的有力工具。然而，近期提出的半监督学习方法数量较少。因此，对半监督学习方法的研究需要更多关注。此外，要注意药物-靶点对涉及结合亲和力和剂量依赖性这一事实。针对药物靶点相互作用预测问题研究新的回归方法更具实际意义。使用定量生物活性数据将带来更准确和可靠的预测结果。最后，随着高通量生物技术的发展，可用的

近来数据增长迅速。是时候进一步利用机器学习技术充分发挥更多不同类型异构数据的作用了。

5. 要点

1. 确定药物与靶点的相互作用是药物研发研究中至关重要的第一步。
2. 现有的多个专业数据库为药物靶点相互作用（DTI）预测提供了已知的数据资源，从而推动了药物研发。
3. 基于机器学习的方法通常对于弥散张量成像（DTI）预测是有效且可靠的。
4. 不同的机器学习方法各有优缺点。因此，为特定的预测任务选择合适的方法或组合模型至关重要。
5. 通过整合更多药物和靶点的异质性数据源，可以建立一个更有效的预测模型。
6. 实际上，药物毒性指数（DTI）预测是一个具有定量生物活性数据的回归问题。

作者贡献：概念化，R.C.；撰写原始草稿，R.C.；撰写、审阅与编辑，R.C.、X.L.、S.J. 和 J.L.（林佳伟）；资金获取，X.L.；监督，J.L.（刘娟）。

资助：本研究得到了国家自然科学基金（项目编号：61472333、61772441、61472335、61425002）、厦门市海洋经济创新发展项目（项目编号：16PFW034SF02）、福建省高等学校自然科学研究项目（项目编号：JZ160400）、福建省自然科学基金（项目编号：2017J01099）、厦门大学校长基金（项目编号：20720170054）以及国家自然科学基金（项目编号：81300632）的资助。

致谢：我们想感谢所有被引用文献的作者。

利益冲突：作者声明不存在利益冲突。

参考文献

1. Masoudi-Nejad, A.; Mousavian, Z.; Bozorgmehr, J.H. Drug-target and disease networks: Polypharmacology in the post-genomic era. *In Silico Pharmacol.* **2013**, *1*, 17. [CrossRef] [PubMed]
2. Paul, S.M.; Mytelka, D.S.; Dunwiddie, C.T.; Persinger, C.C.; Munos, B.H.; Lindborg, S.R.; Schacht, A.L. How to improve R&D productivity: The pharmaceutical industry' s grand challenge. *Nat. Rev. Drug Discov.* **2010**, *9*, 203–214. [CrossRef] [PubMed]
3. Dickson, M.; Gagnon, J.P. Key factors in the rising cost of new drug discovery and development. *Nat. Rev. Drug Discov.* **2004**, *3*, 417–429. [CrossRef] [PubMed]
4. Wang, Y.; Bryant, S.H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B.A.; Thiessen, P.A.; He, S.; Zhang, J. Pubchem bioassay: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D955–D963. [CrossRef] [PubMed]
5. Chen, H.; Zhang, Z. A semi-supervised method for drug-target interaction prediction with consistency in networks. *PLoS ONE* **2013**, *8*, e62975. [CrossRef] [PubMed]
6. Li, J.; Zheng, S.; Chen, B.; Butte, A.J.; Swamidass, S.J.; Lu, Z. A survey of current trends in computational drug repositioning. *Brief. Bioinform.* **2016**, *17*, 2–12. [CrossRef] [PubMed]
7. Zeng, X.; Liu, L.; Lu, L.; Zou, Q. Prediction of potential disease-associated micrornas using structural perturbation method. *Bioinformatics* **2018**, *34*, 2425–2432. [CrossRef] [PubMed]
8. Zhang, X.; Zou, Q.; Rodríguez-Patón, A.; Zeng, X. Meta-path methods for prioritizing candidate disease mirnas. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**. [CrossRef]
9. Hua, S.; Yun, W.; Zhiqiang, Z.; Zou, Q. A discussion of micrornas in cancers. *Curr. Bioinform.* **2014**, *9*, 453–462. [CrossRef]
10. Zeng, X.; Liao, Y.; Liu, Y.; Zou, Q. Prediction and validation of disease genes using hetesim scores. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2017**, *14*, 687–695. [CrossRef] [PubMed]
11. Zeng, J.; Li, D.; Wu, Y.; Zou, Q.; Liu, X. An empirical study of features fusion techniques for protein-protein interaction prediction. *Curr. Bioinform.* **2016**, *11*, 4–12. [CrossRef]
12. Wang, Z.; Zou, Q.; Jiang, Y.; Ju, Y.; Zeng, X. Review of protein subcellular localization prediction. *Curr. Bioinform.* **2014**, *9*, 331–342. [CrossRef]

13. Keiser, M.J.; Roth, B.L.; Armbruster, B.N.; Ernsberger, P.; Irwin, J.J.; Shoichet, B.K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **2007**, *25*, 197−206. [CrossRef] [PubMed]

14. Arola, L.; Fernandez-Larrea, J.; Blay, M.; Salvado, M.J.; Blade, C.; Ardevol, A.; Vaque, M.; Pujadas, G. Protein-ligand docking: A review of recent advances and future perspectives. *Curr. Pharm. Anal.* **2008**, *4*, 1–19. [CrossRef]

15. Yamanishi, Y. Chemogenomic approaches to infer drug−target interaction networks. In *Data Mining for Systems Biology: Methods and Protocols*; Mamitsuka, H., DeLisi, C., Kanehisa, M., Eds.; Humana Press: Totowa, NJ, USA, 2013; Volume 939, pp. 97−113. ISBN 978-1-62703-107-3.

16. Mousavian, Z.; Masoudi-Nejad, A. Drug-target interaction prediction via chemogenomic space: Learning-based methods. *Expert Opin. Drug Metab. Toxicol.* **2014**, *10*, 1273−1287. [CrossRef] [PubMed]

17. Ding, H.; Takigawa, I.; Mamitsuka, H.; Zhu, S. Similarity-based machine learning methods for predicting drug-target interactions: A brief review. *Brief. Bioinform.* **2014**, *15*, 734−747. [CrossRef] [PubMed]

18. Chen, X.; Yan, C.C.; Zhang, X.; Zhang, X.; Dai, F.; Yin, J.; Zhang, Y. Drug-target interaction prediction: Databases, web servers and computational models. *Brief. Bioinform.* **2016**, *17*, 696−712. [CrossRef] PubMed]

19. Ezzat, A.; Wu, M.; Li, X.L.; Kwoh, C.K. Computational prediction of drug-target interactions using chemogenomic approaches: An empirical survey. *Brief. Bioinform.* **2018**. [CrossRef] [PubMed]

20. Luo, Y.; Zhao, X.; Zhou, J.; Yang, J.; Zhang, Y.; Kuang, W.; Peng, J.; Chen, L.; Zeng, J. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.* **2017**, *8*, 573. [CrossRef] [PubMed]

21. Wan, F.; Hong, L.; Xiao, A.; Jiang, T.; Zeng, J. Neodti: Neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions. *Bioinformatics* **2018**. [CrossRef]

22. Ma, T.; Xiao, C.; Zhou, J.; Wang, F. Drug similarity integration through attentive multi-view graph auto-encoders. *arXiv*, **2018**; arXiv:1804.10850.

23. Zhang, W.; Chen, Y.; Liu, F.; Luo, F.; Tian, G.; Li, X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinform.* **2017**, *18*, 18. [CrossRef] [PubMed]

24. Pahikkala, T.; Airola, A.; Pietila, S.; Shakyawar, S.; Szwajda, A.; Tang, J.; Aittokallio, T. Toward more realistic drug-target interaction predictions. *Brief. Bioinform.* **2015**, *16*, 325−337. [CrossRef] [PubMed]

25. Zeng, X.; Zhang, X.; Zou, Q. Integrative approaches for predicting microrna function and prioritizing disease-related microrna using biological interaction networks. *Brief. Bioinform.* **2016**, *17*, 193−203. [CrossRef] [PubMed]

26. Zou, Q.; Ju, Y.; Li, D. Protein folds prediction with hierarchical structured SVM. *Curr. Proteom.* **2016**, *13*, 79–85. [CrossRef]

27. Wang, X.; Zeng, X.; Ju, Y.; Jiang, Y.; Zhang, Z.; Chen, W. A classification method for microarrays based on diversity. *Curr. Bioinform.* **2016**, *11*, 590−597. [CrossRef]

28. Nagamine, N.; Sakakibara, Y. Statistical prediction of protein chemical interactions based on chemical structure and mass spectrometry data. *Bioinformatics* **2007**, *23*, 2004−2012. [CrossRef] [PubMed]

29. Kanehisa, M.; Furumichi, M.; Mao, T.; Sato, Y.; Morishima, K. Kegg: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353−D361. [CrossRef] [PubMed]

30. Placzek, S.; Schomburg, I.; Chang, A.; Jeske, L.; Ulbrich, M.; Tillack, J.; Schomburg, D. Brenda in 2017: New perspectives and new tools in brenda. *Nucleic Acids Res.* **2017**, *45*, D380−D388. [CrossRef] [PubMed]

31. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A. Pubchem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202−D1213. [CrossRef] [PubMed]

32. Qin, C.; Zhang, C.; Zhu, F.; Xu, F.; Chen, S.Y.; Zhang, P.; Li, Y.H.; Yang, S.Y.; Wei, Y.Q.; Tao, L. Therapeutic target database update 2014: A resource for targeted therapeutics. *Nucleic Acids Res.* **2014**, *42*, D1118−D1123. [CrossRef] [PubMed]

33. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z. Drugbank 5.0: A major update to the drugbank database for 2018. *Nucleic Acids Res.* **2017**, *46*, D1074−D1082. [CrossRef] [PubMed]

34. Hecker, N.; Ahmed, J.; Von, E.J.; Dunkel, M.; Macha, K.; Eckert, A.; Gilson, M.K.; Bourne, P.E.; Preissner, R. Supertarget goes quantitative: Update on drug-target interactions. *Nucleic Acids Res.* **2012**, *40*, D1113 D1117. [CrossRef] [PubMed]

35. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; Mcglinchey, S.; Michalovich, D.; Allazikani, B. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107. [CrossRef] [PubMed]

36. Szklarczyk, D.; Santos, A.; Von, M.C.; Jensen, L.J.; Bork, P.; Kuhn, M. STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **2016**, *44*, D380−D384. [CrossRef] [PubMed]

37. Günther, S.; Kuhn, M.; Dunkel, M.; Campillos, M.; Senger, C.; Petsalaki, E.; Ahmed, J.; Urdiales, E.G.; Gewiess, A.; Jensen, L.J. Supertarget and matador: Resources for exploring drug-target relationships. *Nucleic Acids Res.* **2008**, *36*, D919−D922. [CrossRef] [PubMed]

38. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. Bindingdb: A web-accessible database of experimentally determined protein−ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198−D201. [CrossRef] [PubMed]

39. Magariños, M.P.; Carmona, S.J.; Crowther, G.J.; Ralph, S.A.; Roos, D.S.; Shanmugam, D.; Voorhis, W.C.V.; Agüero, F. TDR targets: A chemogenomics resource for neglected diseases. *Nucleic Acids Res.* **2012**, *40*, D1118−D1127. [CrossRef] [PubMed]

40. Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L.J.; Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **2010**, *6*, 343−348. [CrossRef] [PubMed]

41. Seiler, K.P.; George, G.A.; Happ, M.P.; Bodycombe, N.E.; Carrinski, H.A.; Norton, S.; Brudz, S.; Sullivan, J.P.; Muhlich, J.; Serrano, M. Chembank: A small-molecule screening and cheminformatics resource database. *Nucleic Acids Res.* **2008**, *36*, D351−D359. [CrossRef] [PubMed]

42. Liu, Y.; Wei, Q.; Yu, G.; Gai, W.; Li, Y.; Chen, X. DCDB 2.0: A major update of the drug combination database. *Database* **2014**, *2014*. [CrossRef] [PubMed]

43. Kumar, R.; Chaudhary, K.; Gupta, S.; Singh, H.; Kumar, S.; Gautam, A.; Kapoor, P.; Raghava, G.P.S. CancerDR: Cancer drug resistance database. *Sci. Rep.* **2013**, *3*, 1445. [CrossRef] [PubMed]

44. Chen, X.; Ren, B.; Chen, M.; Liu, M.X.; Ren, W.; Wang, Q.X.; Zhang, L.X.; Yan, G.Y. ASDCD: Antifungal synergistic drug combination database. *PLoS ONE* **2014**, *9*, e86499. [CrossRef] [PubMed]

45. Nickel, J.; Gohlke, B.O.; Erehman, J.; Banerjee, P.; Rong, W.W.; Goede, A.; Dunkel, M.; Preissner, R. SuperPred: Update on drug classification and target prediction. *Nucleic Acids Res.* **2014**, *42*, W26−W31. [CrossRef] [PubMed]

46. Mei, J.P.; Kwoh, C.K.; Yang, P.; Li, X.L.; Zheng, J. Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics* **2013**, *29*, 238−245. [CrossRef] [PubMed]

47. Van Laarhoven, T.; Marchiori, E. Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS ONE* **2013**, *8*, e66952. [CrossRef] [PubMed]

48. Yamanishi, Y.; Araki, M.; Gutteridge, A.; Honda, W.; Kanehisa, M. Prediction of drug−target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* **2008**, *24*, i232−i240. [CrossRef] [PubMed]

49. Bleakley, K.; Yamanishi, Y. Supervised prediction of drug−target interactions using bipartite local models. *Bioinformatics* **2009**, *25*, 2397−2403. [CrossRef] [PubMed]

50. Zhang, W.; Zou, H.; Luo, L.; Liu, Q.; Wu, W.; Xiao, W. Predicting potential side effects of drugs by recommender methods and ensemble learning. *Neurocomputing* **2016**, *173*, 979−987. [CrossRef]

51. Shi, J.Y.; Yiu, S.M. SRP: A concise non-parametric similarity-rank-based model for predicting drug-target interactions. In Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC, USA, 9−12 November 2015; IEEE: New York, NY, USA, 2015; pp. 1636−1641.

52. Zong, N.; Kim, H.; Ngo, V.; Harismendy, O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics* **2017**, *33*, 2337−2344. [CrossRef] [PubMed]

53. Cheng, F.; Liu, C.; Jiang, J.; Lu, W.; Li, W.; Liu, G.; Zhou, W.; Huang, J.; Tang, Y. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.* **2012**, *8*, e1002503. [CrossRef] [PubMed]

54. Luo, H.; Wang, J.; Li, M.; Luo, J.; Peng, X.; Wu, F.X.; Pan, Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics* **2016**, *32*, 2664−2671. [CrossRef] [PubMed]

55. Van Laarhoven, T.; Nabuurs, S.B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug−target interaction. *Bioinformatics* **2011**, *27*, 3036−3043. [CrossRef] [PubMed]

56. Gönen, M. Predicting drug−target interactions from chemical and genomic kernels using bayesian matrix factorization. *Bioinformatics* **2012**, *28*, 2304−2310. [CrossRef] [PubMed]

57. Cobanoglu, M.C.; Liu, C.; Hu, F.; Oltvai, Z.N.; Bahar, I. Predicting drug−target interactions using probabilistic matrix factorization. *J. Chem. Inf. Model.* **2013**, *53*, 3399−3409. [CrossRef] [PubMed]
58. Zheng, X.; Ding, H.; Mamitsuka, H.; Zhu, S. Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11−14 August 2013; ACM: New York, NY, USA, 2013; pp. 1025−1033.
59. Ezzat, A.; Zhao, P.; Wu, M.; Li, X.L.; Kwoh, C.K. Drug-target interaction prediction with graph regularized matrix factorization. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *14*, 646−656. [CrossRef] [PubMed]
60. Tenenbaum, J.B.; Silva, V.D.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319−2323. [CrossRef] [PubMed]
61. Yu, H.; Chen, J.; Xu, X.; Li, Y.; Zhao, H.; Fang, Y.; Li, X.; Zhou, W.; Wang, W.; Wang, Y. A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. *PLoS ONE* **2012**, *7*, e37608. [CrossRef] [PubMed]
62. Wang, Y.; Zeng, J. Predicting drug-target interactions using restricted boltzmann machines. *Bioinformatics* **2013**, *29*, i126−i134. [CrossRef] [PubMed]
63. Fu, G.; Ding, Y.; Seal, A.; Chen, B.; Sun, Y.; Bolton, E. Predicting drug target interactions using meta-path-based semantic network analysis. *BMC Bioinform.* **2016**, *17*, 160. [CrossRef] [PubMed]
64. Xia, Z.; Wu, L.Y.; Zhou, X.; Wong, S.T. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst. Biol.* **2010**, *4*, S6. [CrossRef] [PubMed]
65. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans Evol. Comput.* **1997**, *1*, 67−82. [CrossRef]
66. Zhang, P.; Wang, F.; Hu, J.; Sorrentino, R. Label propagation prediction of drug-drug interactions based on clinical side effects. *Sci. Rep.* **2015**, *5*, 12339. [CrossRef] [PubMed]
67. Ezzat, A.; Wu, M.; Li, X.L.; Kwoh, C.K. Drug-target interaction prediction using ensemble learning and dimensionality reduction. *Methods* **2017**, *129*, 81−88. [CrossRef] [PubMed]
68. Ezzat, A.; Wu, M.; Li, X.L.; Kwoh, C.K. Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinform.* **2016**, *17*, 267−276. [CrossRef] [PubMed]
69. Zhang, R. An ensemble learning approach for improving drug−target interactions prediction. In Proceedings of the 4th International Conference on Computer Engineering and Networks, Shanghai, China, 19−20 July 2015; Wong, W.E., Ed.; Springer International Publishing: Cham, Switzerland, 2015; pp. 433−442.
70. Camacho, D.M.; Collins, K.M.; Powers, R.K.; Costello, J.C.; Collins, J.J. Next-generation machine learning for biological networks. *Cell* **2018**, *173*, 1581−1592. [CrossRef] [PubMed]
71. Zeng, X.; Lin, W.; Guo, M.; Zou, Q. A comprehensive overview and evaluation of circular rna detection tools. *PLoS Comput. Biol.* **2017**, *13*, e1005420. [CrossRef] [PubMed]
72. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning (ICML ’ 06), Pittsburgh, PA, USA, 25−29 June 2006; ACM Press: New York, NY, USA, 2006; pp. 233−240.
73. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861−874. [CrossRef]