DOI:10.12113/202409006

基于机器学习的 miRNA 靶标预测方法及相关数据库研究进展

蒋 辉*,罗思杰
(南华大学 计算机学院,湖南 衡阳 421001)

摘 要:微小 RNA(miRNA)通过与靶标 RNA 非翻译区的特定位点结合来调控基因表达。由于识别 miRNA 靶标的高通量实验方法昂贵且耗时,因此,研究可以准确预测 miRNA 靶标的计算方法具有重要意义。本文综述了近年来基于机器学习的 miRNA 靶标预测方法及 miRNA 靶标相关数据库。首先介绍了 miRNA 及其功能,阐明了 miRNA 靶标预测的重要性。随后,概述了常见的 miRNA 靶标数据库,这些数据库为 miRNA 靶标预测提供了重要的数据基础。接着,详细阐述了基于 SVM、集成学习、深度学习等机器学习模型的 miRNA 靶标预测方法。最后,讨论了 miRNA 靶标预测研究的未来挑战和方向,并展望了深度学习技术在 miRNA 靶标预测领域的应用前景。

关键词:miRNA;机器学习;深度学习;miRNA 靶标预测;miRNA 相关靶标数据库
中图分类号:Q522 **文献标志码**:A **文章编号**:1672-5565(2026)01-001-13

Recent development of machine learning-based miRNA target prediction methods and related databases

JIANG Hui*, LUO Sijie
(*School of Computer Science*, *University of South China*, *Hengyang* 421001, *Hunan*, *China*)

Abstract: MicroRNAs (miRNAs) regulate gene expression by binding to specific sites in the non-coding regions of target RNA. Due to the high-throughput experimental methods to identify miRNA targets are expensive and time-consuming, the development of computational methods that can accurately predict miRNA targets is of great significance. In this paper, we reviewed the methods of miRNA target prediction based on machine learning and miRNA target related databases in recent years. First, we introduced miRNAs and their functions, elucidating the importance of miRNA target prediction. After that, we provided an overview of common miRNA target databases, which provide an essential data for miRNA target prediction. Next, we elaborated the miRNA target prediction methods based on SVM, ensemble learning and deep learning. Finally, we discussed the future challenges and research directions on miRNA target prediction, as well as the potential application of deep learning technology in the field of miRNA target prediction.

Keywords: miRNA; Machine learning; Deep learning; miRNA target prediction; miRNA target database

microRNA(miRNA)是一类能够调节基因表达的内源性、非编码、单链小 RNA 分子。它通过与 mRNA 的3'非翻译区(3'UTR)上的特定位点绑定来行使基因调控、抑制基因表达的功能[1]。一般认

收稿日期:2024-09-05;**修回日期**:2025-01-17;**网络首发日期**:2025-05-16.
网络出版地址:https://link.cnki.net/urlid/23.1513.Q.20250515.1727.002
基金项目:湖南省教育厅科学研究项目(No.24A0299);南华大学博士科研启动基金(No.220XQD048).
*通信作者:蒋辉,男,讲师,博士研究生,CCF 会员,研究方向:生物信息与计算生物、智能系统.E-mail:jianghui@usc.edu.cn
引用格式:蒋辉,罗思杰.基于机器学习的 miRNA 靶标预测方法及相关数据库研究进展[J].生物信息学,2026,24(1):1-13.
JIANG Hui, LUO Sijie. Recent development of machine learning-based miRNA target prediction methods and related databases[J]. Chinese Journal of Bioinformatics, 2026, 24(1): 1-13.

为在动物体内,miRNA 与其靶标绑定位点是非完全碱基配对的,这些绑定位点根据规则可分为典型的绑定位点、非典型的绑定位点以及非规范的绑定位点三种类型[2]。如图 1 所示,在 miRNA 序列中,存在一个高度保守的区域,通常位于 miRNA 序列的 5'端第 2-8 个碱基,这个区域被称为 miRNA 的"种子区域"或"种子序列"。典型的绑定位点、非典型的绑定位点要求种子区域与靶标的 3' UTR 完全配对。

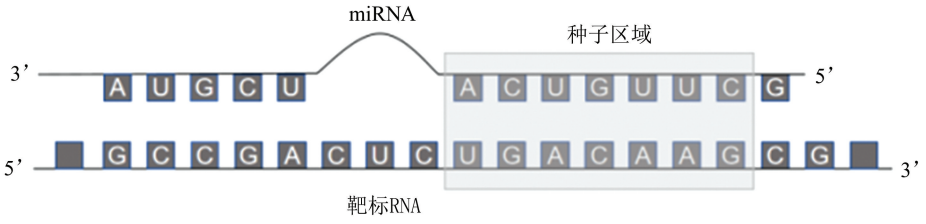

图 1 miRNA 与靶基因绑定示意图

Fig.1 miRNA and target gene binding diagram

miRNA 的异常表达与多种生物过程息息相关,例如肿瘤的抑制和进展[3],神经系统的发育与功能等[4]。因此能精确地识别 miRNA 靶标,对研究人类疾病的产生、发展及治疗有着重大的意义。

但是,由于 miRNA 靶标丰富且广泛存在于转录组当中,所以大量 miRNA 靶标识别具有挑战性。生物学实验方法是识别 miRNA 靶标的经典方法,例如交联免疫沉淀(CLIP)[5]、荧光素酶报告基因实验(Luciferase reporter assays)[6]等。这类方法能够揭示 miRNA 的调控功能和机制,为进一步研究提供实验基础和支持。而受技术限制的这类方法成本高昂、耗时耗力。在研究 miRNA 功能的生物学实验过程中,研究者发现了一些未经充分论证的 miRNA 与其靶标绑定的规则[7],根据这些规则,生物信息学学者们提出了计算方法来预测 miRNA 靶标[8],很好地弥补了生物学实验验证方法的不足。

随着人们对 miRNA 功能研究的不断深入,越来越多的 miRNA 靶标被生物学实验识别出来,出现了一些 miRNA 靶标数据库。这些数据的积累使得机器学习方法可以引入 miRNA 靶标预测的计算方法,进一步提高方法的预测性能。

目前,miRNA 靶标预测计算方法的任务分为两大类:

1)靶位点的预测:该类任务是给定 miRNA 序列和 mRNA 的 3' UTR 序列,从 3' UTR 序列上识别可能的 miRNA 结合位点,以确定是否存在靶向关系。靶位点预测的关键在于提取已知 miRNA 靶位点的序列、结构等特征,并训练分类器来识别 miRNA 靶位点,以确定 miRNA 与 mRNA 的绑定关系。

2)直接预测 miRNA 与基因的绑定关系:该类任务直接预测 miRNA 与基因的表达调控的关系。这种方法的优点在于从整体考虑 miRNA 与靶标之间的复杂相互作用,能够全面分析二者之间的联系。

本文系统地综述了近几年来提出的基于机器学习的 miRNA 靶标预测方法以及 miRNA 靶标数据库。对于 miRNA 靶标数据库,我们根据其所记录数据的类型进行分类汇总。对于 miRNA 靶标预测方法,我们根据所应用的模型进行分类,并对每一类方法进行了分析。最后,我们对这些方法进行了总结,并对未来的研究方向进行了展望。

1 miRNA 靶标数据库

miRNA 靶标的相关数据广泛分布于已发表的科学研究文献中,研究人员采取了人工检索、文本挖掘[9]等方法,将这些 miRNA 靶标信息集中起来,构建成 miRNA 靶标数据库。根据数据库所包含的数据内容,可划分为两大类:一类只记录实验验证的 miRNA 靶标相互作用数据(MTIs)(表 1);另一类则记录了实验验证以及计算方法预测出的 MTIs(表 2)。

1)表 1 给出了 4 种记录实验验证的 MTIs 数据库的相关信息。TarBase v9.0[10]和miRTarBase v9.0[11]都使用自然语言处理(NLP)技术从 PubMed 文献中提取实验验证 MTIs,并对挖掘的数据进行人工审核。其中,TarBase v9.0 相较于 miRTarBase v9.0 收录了数量更多的 MTIs,但是,miRTarBase v9.0 涵盖了更多的物种信息。miRATBase[12]的开发者开发了 HiTmIR 高通量技术来验证 MTIs,并用该数据库存储经过 HiTmIR 验证的 MTIs。

2)表 2 给出了 2 种包含实验验证及计算方法预测的 MTIs 的数据库。miRPathDB[13]数据库整合了

来自 miRBase 数据库(版本 22.1)[14] 和 miRCarta 数据库(版本 1.1)[15] 中所有有关智人(Homo sapiens)和小鼠(Mus musculus)的 miRNA 信息,还记录了通过 TargetScan[16] 和 miRanda[17] 等工具预测的 miRNA 靶标。miRWalk[18] 基于数据库的受欢迎程度和交互预测的准确性整合了 miRTarBase、TargetScan 等数据库中的信息。

还有一类数据库并不包含实验验证的数据,而是专注于汇总 miRNA 靶标预测工具的预测数据(表 3)。由于预测工具的性能依赖于训练数据的质量、所使用的算法和模型,这类数据往往具有假阳性率高等不足之处,但它们为后续研究提供了辅助信息,有助于促进潜在靶点的发现。例如,MirDIP[19] 数据库整合了 24 个预测工具的数据,提供了组织上下文注释、新 miRNA 整理以及上下文特异性分析等功能。

表 1　记录生物学实验验证的 miRNA 靶标的数据库

Table 1　Databases recording miRNA targets verified by biological experiments

年份	名称	记录数	简介	链接
2020	miRATBase[12]	包含 7 个 miRNAs(hsa-miR-129-2-3p、miR-129-3p、miR-873-5p 等)301 个靶标的 636 个 MTIs	miRATBase 是一个收集使用 HiTmIR 技术鉴定的 miRNA 靶标的数据库。它既包含了有关 miRNA 与基因的相互作用信息,也包含了靶位点的信息	https://ccb-web.cs.uni-saarland.de/miratbase (可下载)
2022	miRTarBasev9.0[11]	包含 37 个物种(人、鼠、线虫等)超过 2 200 449个经过验证的 MTIs	miRTarBase 是一个通过文本挖掘方法从相关文献中提取信息,并由人工系统进行整理的 miRNA 靶标数据库。它既包含了有关 miRNA 与基因的相互作用信息,也包含了靶位点的信息	https://miRTarBase.cuhk.edu.cn (可下载)
2023	starBase/ENCORI[20]	包含 23 个物种超过 1 200 000个 CLIP 实验验证的 MTIs 以及超过 459 000 个 degradome 实验验证的 MTIs	starBase/ENCORI 是一个专注于从 CLIP-Seq、degradome-seq 实验数据中挖掘 miRNA 和其它非编码 RNA 的调控网络及其功能的数据库。它既包含了有关 miRNA 与基因的相互作用信息,也包含了靶位点的信息	https://rna.sysu.edu.cn/encori/index.php (可下载)
2024	TarBase v9.0[10]	包含 24 个物种(人、小鼠、秀丽隐杆线虫等)约 600 万个经过验证的 MTIs	TarBase 是一个汇总实验验证(15 种高通量和 22 种低通量实验方法)的 miRNA 靶标数据库。它既包含了有关 miRNA 与基因的相互作用信息,也包含了靶位点的信息	https://dianalab.e-ce.uth.gr/tarbasev9 (可下载)

表 2　记录生物学实验验证及通过计算方法预测的 miRNA 靶标的数据库

Table 2　Databases recording miRNA Targets verified by biological experiments and predicted by computational methods

年份	名称	记录数	简介	链接
2018	miRWalk[18]	实验验证:5 个物种(人类、小鼠、大鼠、牛和狗),共 1 000 292对 miRNA 靶标,计算得到 6 个物种(增加鱼),共 219 577 139 对 miRNA 靶标	miRWalk 是一个开源平台,记录通过实验验证及预测方法(miRDB、TarPmiR、TargetScan 等)预测的 miRNA 靶标数据	https://mpd.bioinf.uni-sb.de (可下载)
2019	miRPathDB v2.0[13]	包含人类的 27 452(候选)个 miRNAs、28 352 个靶标和 16 833 个通路的潜在关联。包含小鼠的 1 978 个 miRNA、24 898 个靶标和 6 511 个功能类别的相互作用信息	miRPathDB 2.0 是一个专注于汇总实验验证及预测方法(TargetScan、miRanda 等)预测的、有关人类和小鼠 miRNA 靶标的数据库	https://mpd.bioinf.uni-sb.de (可下载)

表 3 记录计算方法预测的 miRNA 靶标的数据库

Table 3 Databases recording miRNA targets predicted by computational method

年份	名称	记录数	简介	链接
2019	miRDB 2020[21]	包括五种物种(人类、小鼠、大鼠、狗和鸡)中受 7 000多个 miRNAs 调控的 350 万个预测靶基因	miRDB 是一个在线数据库,专门用于预测 miRNA 靶基因以及进行功能注释。该数据库中的所有靶基因都是通过生物信息学工具 MirTarget 预测获得的,该工具是基于对数千个 miRNA 与靶标相互作用的高通量测序实验结果的分析而开发的	http://mirdb.org (可下载)
2022	MirDIP v5.2[19]	包含 2 734 个 miRNAs 和 27 936个基因的 46 364 047 个预测 MTIs	mirDIP 是一个记录计算方法预测的 miRNA 靶标相互作用的数据库,主要收集和整合来自多个数据库(TargetScan v7.2、RNA22、rnahybrid May 2021 等)的 miRNA 与靶标的相互作用预测信息,并对其进行注释。数据主要来源包括 miRNA 预测数据库、基因表达数据库以及文献报道	https://ophid.utoronto.ca/mirDIP (可下载)

2 基于机器学习的 miRNA 靶标预测方法进展

随着生物学实验不断发现新的 miRNA 靶标,以及 miRNA 靶标数据库的建立,为 miRNA 靶标预测的计算方法带来了新的发展机遇。鉴于基于机器学习的 miRNA 靶标预测方法成为研究热点,我们整理了近几年出现的相关方法,并根据它们所使用的模型分为四大类:基于支持向量机(Support vector machine, SVM)的方法、基于集成学习的方法、基于深度学习的方法及其他类型的方法,如图 2 所示。

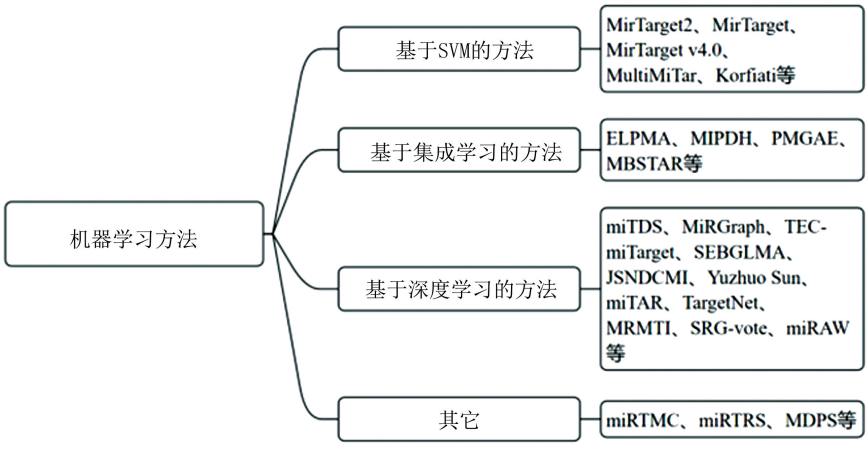

图 2 miRNA 靶标预测方法的分类图

Fig.2 Classification of miRNA target prediction

这些方法大多采用监督学习策略,首先构建数据集,其正样本来源于 miRNA 靶标数据库中实验验证的 miRNA 靶标对,负样本则来源于以下途径:①miRNA 靶标数据库中的少量负样本;②设定规则从生物学实验数据中抽取;③通过计算方法从未标记样本中选取。随后,从这些正负样本中提取序列特征和/或结构特征,用于训练和优化 miRNA 靶标预测的分类器。最终,经过训练的分类器将用于预测新的 miRNA 靶标候选结合位点或绑定关系。表 4展示了近年来基于机器学习的 miRNA 靶标预测方法的相关信息。

2.1 基于支持向量机(SVM)的方法

miRNA 靶标预测是一个复杂的任务,需要综合考虑多种特征,例如序列互补性、碱基配对能量、miRNA 结合位点周围的保守性等等,这些特征之间可能存在非线性关系,传统的线性模型难以有效处理。SVM 模型可以整合不同类型和性质的特征,例如数值特征和交互特征等,并将它们纳入同一个框

架进行分析;SVM 还可以将原始特征映射到高维空间,从而有效地处理非线性问题。

Wang 等[22]开发的 MirTarget2 方法从 Linsley 等[23]发表的 miRNA 过表达微阵列数据中识别出表达水平显著降低的基因,并从 NCBI GEO[24]数据库中选取下调基因所对应的 3' UTR 序列组合成 miRNA 靶标对作为正样本,表达水平没有显著变化的基因对应的 3' UTR 序列组合成 miRNA 靶标对作为负样本。Wang 等[22]从 3' UTR 序列中提取了种子序列保守性、碱基组成、二级结构稳定性等 131 个特征组合成特征向量输入到 SVM 进行训练预测。在 10 折交叉试验中,该方法的 ROC 曲线下面积(AUC)为 0.79。

2016 年,Wang 等[25]开发了名为 MirTarget 的新型 miRNA 靶标预测方法。该方法从 Grosswendt[26]和 CLASH[27]数据集中筛选出 miRNA 种子序列与靶标完全配对的序列对作为正样本,通过随机选取非靶标序列对作为负样本。该方法提取了种子位点保守性、靶标在 UTR 中的位置及靶标的 GC 含量等 50 个特征来训练 SVM。10 折交叉验证结果显示,方法的 AUC 值达到 0.86,优于对比方法。

2019 年,Liu 等[28]进行了 RNA-Seq 实验,将 miRNA 过表达后显著下调的基因转录本以及与 miRISC 结合的转录本作为正样本,没有显著变化以及在 CLIP 结合数据中没有被结合的转录本作为负样本。Liu 等[28]提取了包含 miRNA 的表达水平、miRNA 的成熟度、miRNA 的亚细胞定位等 96 个特征,利用 SVM 模型进行训练开发了 MirTarget v4.0。在独立集实验结果的 ROC 曲线分析中,该方法的 AUC 值为 0.78;在精确度-召回率曲线分析中,当召回率低于 20%时,该方法的精确率超过 90%。

除 MirTarget 之外,MultiMiTar[29]方法以及 Korfiati 等[30]提出的方法也选择了 SVM 作为分类器来预测 miRNA 与靶标之间的结合位点。其中,MultiMiTar 方法从相关文献[31]中获取实验验证的正样本和负样本,并提取了沃森-克里克配对数量、种子位点上游和下游区域的 AU 富集情况等 90 个特征来训练 SVM 分类器;Korfiati 等提出的方法从 TarBase v5.0 数据库和 miRecords[32]数据库中获取正样本和负样本,并通过随机生成 miRNA 序列后利用 miRanda 程序生成负样本补充,该方法抽取包括 miRNA 靶标的结构特征、热力学特征、位置特征等 124 个特征,并利用遗传算法进行参数优化和最终特征选取,最后使用 SVM 进行训练和预测。

2.2　基于集成学习的方法

基于集成学习方法的 miRNA 靶标预测主要包括以下几个步骤:首先,进行数据预处理以及负样本的筛选。接着,提取 miRNA 和靶标的多种特征,包括序列特征和结构特征等;然后,利用训练集训练多个个体学习器,并采用集成策略(如投票、加权平均等)结合各个个体学习器的预测结果;最终,通过集成多个学习器的优势,形成一个强学习器进行预测。集成学习能够有效地减少单一学习器可能带来的过拟合风险,并提高预测精度,增强模型的稳定性和泛化能力。

MBSTAR[33]方法从 miRecords 数据库中获取经实验验证的 miRNA-mRNA 正样本,筛选在表达水平上没有相互作用且能量和保守评分较差的 miRNA-mRNA 对为负样本。该方法在 mRNA 的 3' UTR 区域识别潜在结合位点(PBS),从每个 PBS 及其上下游区域提取序列和结构特征,并通过无监督的拉普拉斯特征选择方法筛选出 PBS 周围区域的单个核苷酸频率、最小自由能等 40 个最具区分性的特征。随后,利用这些特征训练随机森林分类器以预测 miRNA 靶标。在基于 TarBase v6.0 的独立测试集上,MBSTAR 的 F1-score 达到0.337。相较于对比方法,MBSTAR 在预测靶标和功能结合位点(FBS)方面更强。

TarPmiR[34]方法从 CLASH 实验中获取正样本,在与正样本相同的 mRNA 上选取与正样本具有相似的序列组成,但不与正样本重叠、且具有最低折叠能的序列作为负样本。方法提取了折叠能、AU 含量、靶标区域长度等 13 个特征训练随机森林模型,并通过投票方式对预测结果进行集成。相较于对比方法,TarPmiR 的召回率为 55.1%,高出至少 10%。

MIPDH[35]方法利用 NPInter v2.0[36]、DisGeNET[37]等数据库及来源于 NCBI 网站的数据构建包含药物、疾病和 6 种生物分子之间 17 种关联的异构网络。该方法以实验验证的 miRNA-mRNA 相互作用数据为正样本,并随机生成等量不与正样本重叠负样本。通过 DeepWalk 算法学习网络中每个节点的潜在特征,并结合 K-mer 学习 miRNA 和靶标序列的属性特征,形成最终输入特征向量。方法采用随机森林模型进行训练和预测。5 折交叉验证结果显示,MIPDH 方法的精确率、特异性和 AUC 分别为 76.66±1.00%、77.33±1.34%和0.804 4±0.007 8,展现出良好的预测性能。

PMGAE[38]方法从 starBase v2.0[39]数据库中获取已知的假基因-miRNA 关联对作为正样本,同时随机抽取与正样本数量相等的数据作为负样本。该方法使用 3-mer 将每个 miRNA 序列表示为一个 64(4×4×4)维的向量,使用 Jaccard 相似度、余弦相似度和皮尔逊相似度计算假基因和 miRNA 的相似度特征

并进行加权相似融合。基于特征信息和关联信息,通过图自动编码器(GAE)获取节点低维表示向量,然后将这些表示向量输入到 XGBoost 分类器中,以进行假基因-miRNA 关联预测。在 5 折交叉验证评估模型中,平均 AUC 为0.863 4,平均 AUPR 为0.896 6。

ELPMA[40]方法从 starBase v2.0 数据库中选取经过实验验证的假基因-miRNA 作用作为正样本,通过重采样方法从未经证实的假基因-miRNA 对中随机选取与正样本数量相同的对作为负样本构建多个平衡的假基因-miRNA 关联训练子集。ELPMA 使用相似性核融合方法将假基因和 miRNA 的各种相似度特征(假基因表达相似度、miRNA 功能相似度等)进行融合,得到综合相似度矩阵。利用 XGBoost 模型作为个体学习器进行训练,最后使用软投票策略进行集成。在 5 折交叉验证中,方法的 AUC 和 AUPR 分别达到0.989 7和0.991 4。

JSNDCMI[41]方法从 CircBank 数据库[42]和 CMI-9905 数据集[43]中提取已知的 circRNA-miRNA 相互作用(CMI)对作为正样本,通过随机选取获取负样本。该方法从 CMI 网络中提取了分子的局部拓扑结构特征和功能相似性特征,从 RNA 序列和 RNA 相似性网络中提取分子属性特征,通过将分子属性特征和网络嵌入特征进行融合,形成最终的分子特征描述符矩阵作为输入。该方法使用降噪自编码器(DAE)对矩阵进行特征提取。最后,利用梯度提升决策树(GBDT)分类器进行训练,以预测潜在的 CMI。在 5 折交叉验证中,JSNDCMI 在构建的两个数据集上分别获得了 0.900 3 和 0.943 5 的 AUC 值,均优于对比方法。

SEBGLMA[44]方法从 lncRNASNP2[45]数据库中提取已知的 lncRNA-miRNA 相互作用对作为正样本,通过随机选取的方法获取负样本。该方法将 lncRNA 和 miRNA 序列进行 K-mer 分割,使用 word2vec 模型(Skip-gram)将 K-mer 序列转换为词嵌入向量,分别构建 Lnc2Vec 和 Mi2Vec 模型,利用 GIP 模型计算 lncRNA 和 miRNA 的自相似度,并结合已知交互数据构建完整的邻接矩阵。该方法将词嵌入向量和邻接矩阵输入图卷积网络中学习节点属性特征和行为特征,使用旋转森林对融合特征进行分类,预测 lncRNA-miRNA 交互的可能性。SEBGLMA 在 5 折交叉验证中取得了优异的性能,平均准确率达到 87.09%。与现有预测 lncRNA-miRNA 相互作用的模型相比,SEBGLMA 取得了更高的 AUC 值。

2.3 基于深度学习的方法

随着 RNA-Seq 等技术的快速发展,越来越多的 miRNA 靶标被挖掘出来,这为基于深度学习模型的 miRNA 靶标预测方法提供了数据基础。图 3 给出了基于深度学习方法的 miRNA 靶标预测的主要步骤。这类方法首先收集经实验验证的 miRNA 靶标数据,从中获取正样本和负样本。然而,由于负样本数量稀缺,需要采用算法来扩充负样本集。接着,对数据进行预处理,利用 one-hot 编码等将序列转换为深度学习模型可识别的数值表示;结合 miRNA 靶标相互作用关系以及相似性等来构建异构图。随后,利用 CNN、LSTM、RNN 等深度学习模型从 miRNA 和 mRNA 序列中提取特征;使用 GCN,HGT 等图神经网络从构建的异构图中提取特征。然后,通过全连接层将特征表示转换为预测结果。最后,使用测试集、交叉验证、消融试验等方法对模型进行评估。

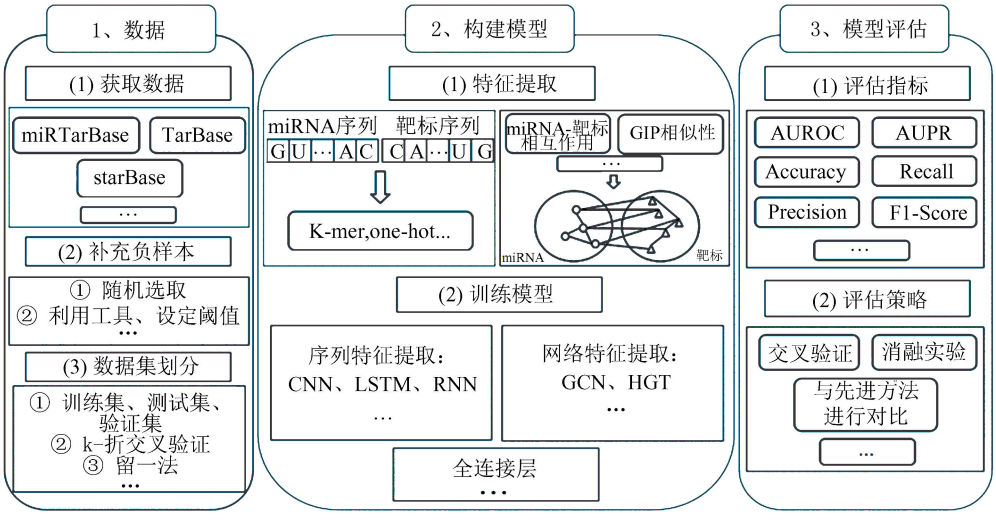

图 3 基于深度学习方法的 miRNA 靶标预测主要步骤

Fig.3 Main steps for miRNA target prediction using deep learning methods

miRAW[46]方法从 TarBase v7.0 和 miRTarBase v6.0 数据库获取正样本和负样本,针对负样本不足的情况,使用 ViennaRNA 包中的工具预测 miRNA 与 mRNA 片段形成的二级结构,并筛选 miRNA 与 mRNA 片段形成双链时、自由能变化小于 0 的潜在靶标作为负样本补充。该方法使用 one-hot 编码对 miRNA 序列以及 3' UTR 序列进行编码;通过由 8 个密集隐藏层和 2 个 softmax 输出节点组成的人工神经网络(ANN)进行预测,其中隐藏层 1~5 层作为自编码器学习 miRNA 与结合位点复合物的特征表示,隐藏层 6~8 层负责将自编码器学习的特征进行分类;最后通过 softmax 函数输出概率分布。在 10 折交叉验证中,miRAW 在准确性和 F1 分数方面均优于对比方法。

miTAR[47]方法采用 DeepMirTar[48]和 miRAW 两个数据集。对于 DeepMirTar 数据集,方法选择了仅包含在 3' UTR 上的 miRNA 靶标作为正样本,并通过将 miRNA 序列随机打乱生成负样本;对于 miRAW 数据集,通过随机选取与训练、测试和验证数据集中未重叠的数据获取独立集。该方法使用嵌入层将 miRNA 和基因序列中的一个碱基编码为 5 维向量进行输入;通过 CNN 学习 miRNA 与靶标之间的空间特征,利用双向 RNN 学习序列特征;最后利用两个全连接层进行输出。该方法与 DeepMirTar 和 miRAW 进行了比较,在两个数据集上,该方法的准确率分别达到 96.49%和 97.87%,表现更优。在构建的独立集上,该方法的准确率达到了 96.6%。

TargetNet[49]方法从 miRAW 数据集中获取正样本,通过使用滑动窗口识别 miRNA-CTS(候选靶位点)对,并使用 RNACofold 工具筛选出二级结构产生负自由能的序列对作为负样本。该方法进行了扩展种子区域的比对,并将比对结果、miRNA 及 CTS 序列进行 one-hot 编码,然后将编码后的序列进行拼接和零填充后作为输入。该方法利用 ResNet 模型进行训练预测,利用全连接层进行输出,并选择一个 miRNA-mRNA 对中所有 miRNA-CTS 对的预测分数的最大值作为该 miRNA-mRNA 对的最终分数。实验结果显示,TargetNet 的 F1 分数和准确率均优于对比方法。

SRG-vote[50]方法使用 miRTarBase v7.0 中的 miRNA-基因对作为正样本,通过距离方法[51]生成负样本。该方法通过计算 miRNA 序列之间的余弦相似度构建了 miRNA-miRNA 相似性网络,通过 BioGrid PPI 网络[52]构建基因-基因网络,在此基础上通过引入 MISIM[53]及 HumanNet[54]数据集构建扩展的 miRNA-miRNA 相似性网络和基因-基因网络。该方法将 miRNA 和基因序列及构建的网络作为模型的输入;使用 doc2vec 从 miRNA 和基因序列中提取序列嵌入;使用 role2vec[55]从 miRNA-miRNA 网络和基因-基因网络中提取几何嵌入;使用 GCN 从扩展的 miRNA-miRNA 网络和基因-基因网络中提取几何嵌入,并利用 LSTM 或双向(BiLSTM)模型分别对三种嵌入进行训练,学习 miRNA-基因对之间的潜在关系;最后使用投票系统将三种嵌入的预测分数进行加权平均,得到最终的预测分数。实验显示,SRG-vote 相较于对比方法具有更高的 AUC。

MRMTI[56]方法从 miRTarBase v8.0 中获取正样本,通过选取随机选取与正样本数量相同的、未经实验验证的 miRNA-基因对作为负样本。该方法使用 one-hot 编码 miRNA 和基因序列,利用 word2vec 将基因序列转换为词嵌入后使用 BiLSTM 学习序列特征表示;通过计算 miRNA 序列之间的相似度构建 miRNA 相似度网络;通过计算基因功能相似度构建基因相似度网络;通过实验验证的数据构建 miRNA-基因相互作用网络,并利用多关系图卷积网络进行网络的特征提取;最后将网络嵌入和序列嵌入融合到同一嵌入空间中进行预测。与其他方法相比,MRMTI 在 AUC(0.918 3)和 AUPR(0.920 4)等指标上表现更优。

Sun 等[57]从位点级(Site-level)和基因级(Gene-level)两个层级来考虑 miRNA 靶标预测,其中,从 PAR-CLIP[58]和 CLASH 数据中获取位点级预测的数据集,将具有负自由能的配对作为正对,长度不超过 30 个核苷酸且形成稳定双链的配对视作为负对。从 Tarbase v7.0 和 miRTarBase v8.0 数据库中获取基因集预测的数据集,通过筛选和删除不同实验中结果矛盾的数据,合并重复数据获取实验验证的正对和负对。该方法使用 skip-gram 模型分别对 miRNA 和 mRNA 序列进行训练,得到每个核苷酸的向量表示。将原始的 miRNA 和 CTS 序列中的每个核苷酸替换为对应的向量表示,并将转换后的 miRNA 和 CTS 序列矩阵分别输入到 BiLSTM 模型中进行特征提取,随后将提取的特征进行拼接后利用第二层 BiLSTM 进行更复杂的特征学习;最后利用 softmax 函数进行输出。与数据库预测工具(PITA 和 TargetScan)和深度学习方法(CNN1d 和 DeepMirTar)相比,该方法具有更高的准确性、特异性、ROC 值等。

DMISO[59]方法将 CLASH 数据中提取的 miRNA/isomiR 及其对应的 mRNA 靶标位点序列相互作用对作为正样本,对于每个已知的 miRNA/isomiR 正样本,在对应 mRNA 转录本的 3' UTR 中选

择一个低自由折叠能量的位点作为负样本。该方法将 miRNA/isomiR 序列和 mRNA 靶标位点序列分别进行 one-hot 编码;利用 CNN 提取 miRNA/isomiR 和靶标位点的空间特征,并利用双向 LSTM 将提取的特征进行整合;最后通过全连接层输出,预测 miRNA/isomiR 与靶标之间是否相互作用。在 10 折交叉验证中,DMISO 表现出高精度(95%)和高召回率(90%)。

TEC-miTarget[60] 方法从 deepTargetPro[61]、DeepMirTar 和 miRAW 中获取数据,其中,deepTargetPro 数据集包含 miRNA-mRNA 序列对,miRAW 和 DeepMirTar 数据集包含 miRNA-CTS 序列对。miRAW 和 deepTargetPro 的正负样本均通过实验方法确定。DeepMirTar 中的正样本通过实验方法确定,负样本通过随机改变 miRNA 序列的碱基顺序,选取与靶基因序列无法有效结合的序列生成。该方法使用嵌入层对四种碱基和零填充进行编码,使用位置编码器对碱基的位置信息进行编码,然后将两个向量的和作为输入。该方法使用 Transformer 来捕获 miRNA 及其 CTS 序列的特征表示,然后将这些特征进行融合;利用 CNN 进行特征提取和预测;通过 sigmoid 函数进行输出,判断预测的 CTS 序列是否是 miRNA 靶标。在序列水平上,TEC-miTarget 与 miRAW 和 DeepMirTar 进行比较,结果显示,TEC-miTarge 在准确度、F1 分数、AUC 等指标上表现更优。在转录本水平上,TEC-miTarget 与 deepTargetPro、PITA、mirSVR 等方法进行了比较,虽然 miRDB 在特异性方面表现较好,但 TEC-miTarget 在其他性能指标上表现更优。

MiRGraph[62] 方法通过融合网络信息和序列信息来提升 miRNA 靶点预测的准确性。方法使用 TargetScan 预测的 miRNA-基因相互作用网络为基础,使用 miRTarBase 数据库中的 miRNA-基因相互作用来标记 TargetScan 预测的边,将其分为正样本和负样本。方法构建了三个相互作用网络,基于 miRNA 家族信息[63] 构建的 miRNA-miRNA 相互作用网络,基于 STRING 数据库[64] 构建地基因-基因相互作用网络,以及基于 TargetScan 构建的 miRNA-基因相互作用网络。该方法利用 HGT 模块[65] 捕捉网络的拓扑特征,并通过基于 Transformer 的 CNN 模块分别提取 miRNA 和基因的序列特征,通过 MLP 层将网络特征和序列特征进行整合,并使用双线性函数计算预测分数,利用 Sigmoid 函数进行输出。实验结果显示,MiRGraph 方法在 AUROC 和 AUPR 等指标上均优于对比方法。

miTDS[66] 方法从 TarBase v7.0 和mirTarbase v6.0数据库中获取正样本,通过随机生成与 mRNA 序列长度相同但碱基组成不同的序列构建负样本。miTDS 包含两种输入:第一种是使用 BERT 模型对 miRNA 序列和 CTS 序列进行编码后得到的每个序列的动态语义特征向量;第二种是使用 Biopython 的 pairwise2 包[67] 对 miRNA 的扩展种子区域与 CTS 序列进行全局对齐后得到的扩展种子区域的对齐特征向量。该方法将两种特征进行拼接后利用多尺度 inception 模块[68]、BiLSTM 进行特征学习,最后将学习到的特征输入到 MLP 中,得到每个 CTS 序列的预测分数。miTDS 在召回率、精确率和 F1 分数等指标上均优于对比方法。

2.4 其它

miRTRS[69] 是一种基于推荐模型、使用二分图推理的预测方法。方法利用已知 miRNA 靶标数据,构建 miRNA-基因关系二分图,通过二分图推理的方法来预测关联边。为了精确的预测 miRNA 靶标,该方法利用 miRNA 种子区域序列相似度,结合已知的 miRNA 或基因交互数据构建交互网络,并构建加权有向图,计算基因之间的相似度。最终,通过计算 miRNA 与基因之间的推荐分数,预测潜在的 miRNA 靶标。为了缓解“冷启动”问题,方法设计算法为新的 miRNA-基因预先给出关联边。并在 AUC 等指标上优于对比的 miRNA 靶标预测方法。

miRTMC[70] 方法从 miRTarBase 版本 6.1 以及版本 7 中获取实验验证的 miRNA-基因相互作用数据,构建了 miRNA 相似度网络、基因相似度网络和 miRNA-基因相互作用网络,并将三个网络进行整合,构建 miRNA-基因异构网络。该方法假设异构网络的邻接矩阵是低秩矩阵,利用矩阵补全的方法来预测 miRNA 靶标,在求解过程中,通过 ADMM 算法迭代更新目标矩阵和辅助变量,并利用 R4SVD 算法加速核范数的计算,最终得到补全后的矩阵。该方法构建了两个数据集,并在其上进行了 10 折交叉验证,实验结果表明,miRTMC 要优于对比方法。

Amlan 等[71] 开发了一种名为 MDPS 的算法,用于预测 miRNA 靶标预测。研究人员将 miRNA 与靶标之间的配对模式分为五种状态:匹配、错配、G-U 匹配、miRNA 凸出、靶标突出,利用马尔可夫模型的权重矩阵和转移矩阵计算 miRNA 与靶标之间配对的得分,得分越高,表示配对越稳定,预测为 miRNA 靶标的可能性越大。研究人员将 MDPS 算法与 miRanda 等 miRNA 靶标预测工具相结合,发现引入该算法可以进一步提高预测的精确度。

表 4　预测方法简括表

Table 4　Brief description of prediction methods

年份	方法名称	所用模型	实验评估	工具链接
2019	MirTarget v4.0[28]	SVM	在基于 HITS-CLIP 实验的数据集上进行评估: AUC:0.78	http://mirdb.org (可下载)
2020	MDPS[71]	马尔可夫模型	加入 MDPS 打分后:TargetScan 在 CLASH 数据集上预测的 F1 分数和 Precision 提升了 20.78%和 23.37%	http://hulab.ucf.edu/research/projects/MDPS/index.html (可下载)
2020	MIPDH[35]	DeepWalk、RF	在基于 NPInter v2.0、DisGeNET 和 NCBI 的数据集上进行 5 折交叉验证:AUC:0.8044	需要向作者进行申请。
2020	miRTMC[70]	矩阵补全	在基于 miRTarBase v6.1 的数据集上进行 10 折交叉验证:AUC:0.9285 在基于 miRTarBase v7 的数据集上进行 10 折交叉验证:AUC:0.9216	https://github.com/hjiangcsu/miRTMC(可下载)
2021	PMGAE[38]	GAE、XGBoost	在基于 starBase v2.0 的数据集上进行 5 折交叉验证: AUC:0.8634 AUPR:0.8966	需要向作者进行申请。
2021	miTAR[47]	CNN、BiRNN	在基于 miRAW[46] 的数据集上进行评估: accuracy:0.9649 recall:0.9616 在基于 DeepMirTar[48] 的数据集上进行评估: accuracy:0.9787 recall:0.9717	https://github.com/tjgu/miTAR(可下载)
2022	TargetNet[49]	ResNet	在基于 miRAW 的数据集上进行评估: accuracy:0.7251 negative precison:0.8966 recall:0.9411	https://github.com/mswzeus/TargetNet(可下载)
2022	SRG-vote[50]	doc2vec、GCN、role2vec、LSTM、BiLSTM	在基于 miRTarBase v7.0 的数据集上进行 5 折交叉验证:AUC:0.958	https://github.com/Xshelton/SRG-vote (可下载)
2022	MRMTI[56]	GCN、word2vec、BiLSTM	在基于 miRTarBase v8.0 的数据集上进行评估: AUC:0.9183 AUPR:0.9204	https://github.com/ojinxj/MRMTI (可下载)
2022	Yuzhuo Sun、Fei Xiong 等[57]	BiLSTM	在基于 PAR-CLIP 和 CLASH 的数据集上进行 5 折交叉验证: AUC:0.9773 AUPR:0.9777	需要向作者进行申请。
2022	DMISO[59]	CNN、BiLSTM	在基于 CLASH 的数据集上进行 10 折交叉验证: AUC:0.9945 AUPR:0.9969	https://github.com/qi29(可下载)
2023	ELPMA[40]	SKF、XGBoost	在基于 starBase v2.0 的数据集上进行 5 折交叉验证: AUC:0.9897 AUPR:0.9914	需要向作者进行申请。
2023	JSNDCMI[41]	Node2vec、GBDT、truc2vec、DAE、	在基于 CMI-9905 的数据集上进行 5 折交叉验证: AUC:0.9003 AUPR:0.8999	需要向作者进行申请。

续表 4

年份	方法名称	所用模型	实验评估	工具链接
2023	SEBGLMA[44]	word2vec、GIP、GCN、ROF	在基于 lncRNASNP2 的数据集上进行 5 折交叉验证： AUC：0.9301 AUPR：0.9323	https://github.com/zhaozhiya-20/SEBGLMA-semantic-embedded-bipartite-graph-network-for-predicting-lncRNA-miRNA-associations（可下载）
2023	TEC-miTarget[60]	Transformer Encoder、CNN	在基于 miRAW 的数据集上进行评估： AUC：0.9936 AUPR：0.9941 在基于 DeepMirTar 的数据集上进行评估： AUC：0.9873 AUPR：0.9918	https://github.com/tingpeng17/TEC-miTarget（可下载）
2023	MiRGraph[62]	HGT、CNN、Transformer	在基于 miRTarBase 和 TargetScan 的数据集上进行评估： AUC：0.8706 AUPR：0.8109	https://github.com/Liangyushi/MiRGraph（可下载）
2024	miTDS[66]	Transformer	在基于 miRAW 的数据集上进行评估： accuracy：0.77 negative precison：0.9326 recall：0.9578	https://github.com/mingziiz/miTDS（可下载）

3 总结与展望

研究 miRNA 靶标对于揭示 miRNA 在生物学中的功能，尤其是在人类疾病的发生、发展及预后中的作用，具有重要意义。深入理解 miRNA 靶标的作用不仅有助于疾病机制的研究，还能够推动药物开发和个性化医疗的实现。本文回顾了近年来 miRNA 靶标数据库和靶标预测方法的最新进展。

尽管已有多种 miRNA 靶标预测方法被提出，但这一领域仍面临诸多挑战，亟待进一步优化和突破。首先，特征提取比较困难。由于 miRNA 本身的长度较短，使得传统的序列特征提取方法难以捕捉到足够的生物学信息，尽管深度学习方法（如 CNN、LSTM 等）在一定程度上可以缓解这一问题，但这类方法在面临大规模数据的时候比较费时间和空间，效率略显不足。如何有效地从短序列中提取有意义的特征仍是一个挑战。研究表明，每个 miRNA 平均能够靶向超过 100 个基因位点，某些 miRNA 甚至可以靶向数千个基因位点。在此背景下构建的异构网络中，节点与边的特征异常复杂，如何针对每个节点和边进行特征提取，并且有效地融合这些信息，构建出准确的预测模型，是目前利用图神经网络提取特征的一大难点。其次，负样本数据不足。这影响了训练集的构建，导致在模型训练过程中，尤其是监督学习模型的训练效果受到限制，进而影响了模型的预测能力和泛化能力。

面对这些挑战，未来的研究方向可以集中在以下几个方面。首先，结合实验数据的持续积累，利用迭代学习策略进一步推动模型的更新和优化，通过定期引入新的实验数据，不断调整和优化模型，提升预测准确性。同时，期待未来生物学家和计算机科学家之间更紧密的合作，推动数据共享，提高数据质量。其次，在构建 miRNA 靶标结合异构图时尝试整合更丰富的节点和边信息。其次，尝试利用融合策略，将序列特征和网络特征进行结合来提升预测效果。最后，深度学习技术也在持续进步，未来还应探索新方法以进行特征提取，尤其是能够同时建模节点与边特征的技术，以期在 miRNA 靶标预测中取得突破。

参考文献（References）

[1] YE Jiawei, XU Mingcheng, TIAN Xueke, et al. Research advances in the detection of miRNA[J]. Journal of Pharmaceutical Analysis, 2019, 9(4): 217-226. DOI: 10.1016/j.jpha.2019.05.004.

[2] BARTEL, DAVID P. Metazoan micrornas[J]. Cell, 2018, 173(1): 20-51. DOI: 10.1016/j.cell.2018.03.006.

[3] JANG H, PARK S, KIM J, et al. The tumor suppressor, p53, negatively regulates non-canonical NF-κB signaling through miRNA-induced silencing of NF-κB - inducing kinase[J]. Molecules and Cells, 2020, 43(1): 23-33. DOI:

10.14348/molcells.2019.0239.

[4]BRENNAN G P, HENSHALL D C. MicroRNAs as regulators of brain function and targets for treatment of epilepsy[J]. Nature Reviews Neurology, 2020, 16(9): 506-519. DOI: 10.1038/s41582-020-0369-8.

[5]LEE F C Y, ULE J. Advances in CLIP technologies for studies of protein-RNA interactions[J]. Molecular Cell, 2018, 69(3): 354-369. DOI: 10.1016/j.molcel.2018.01.005.

[6]TOMASELLO L, CLUTS L, CROCE C M. Experimental validation of microRNA Targets: Luciferase reporter assay[M]. MicroRNA Target Identification: Methods in Molecular Biology, New York: Humana Press,2019. DOI: 10.1007/978-1-4939-9207-2_17.

[7]WITKOS T M, KOSCIANSKA E, KRZYZOSIAK W J. Practical aspects of microrna target prediction[J]. Current Molecular Medicine, 2011, 11(2): 93-109. DOI: 10.2174/156652411794859250.

[8]FEITOSA R M M W, PRIETO-OLIVEIRA P, BRENTANI H, et al. MicroRNA target prediction tools for animals: Where we are at and where we are going to-A systematic review[J]. Computational Biology and Chemistry, 2022, 100: 107729. DOI: 10.1016/j.compbiolchem.2022.107729.

[9]KHATUN M S, ALAM M A, SHOOMBU-ATONG W, et al. Recent development of bio-informatics tools for microRNA target prediction[J]. Current Medicinal Chemistry, 2022, 29(5):865-880. DOI:10.2174/0929867328666210804090224.

[10]SKOUFOS G, KAKOULIDIS P, TASTSOGLOU S, et al. TarBase-v9.0 extends experimentally supported miRNA-gene interactions to cell-types and virally encoded miRNAs[J]. Nucleic Acids Research, 2024, 52(D1): D304-D310. DOI: 10.1093/nar/gkad1071.

[11]HUANG Hshiyuan, LIN Yangchidung, CUI Shidong, et al. miRTarBase update 2022: An informative resource for experimentally validated miRNA - target interactions[J]. Nucleic Acids Research, 2022, 50(D1): D222-D230. DOI: 10.1093/nar/gkab1079.

[12]KERN F, KRAMMES L, DANZ K, et al. Validation of human microRNA target pathways enables evaluation of target prediction tools[J]. Nucleic Acids Research,2021,49(1): 127-144. DOI: 10.1093/nar/gkaa1161.

[13]KEHL T, KERN F, BACKES C, et al. miRPathDB 2.0: A novel release of the miRNA pathway dictionary database[J]. Nucleic Acids Research, 2020, 48(D1): D142-D147. DOI: 10.1093/nar/gkz1022.

[14]KOZOMARA A, BIRGAOANU M, GRIFFITHS-JONES S. miRBase: From microRNA sequences to function[J]. Nucleic Acids Research, 2019, 47(D1): D155-D162. DOI: 10.1093/nar/gky1141.

[15]BACKES C, FEHLMANN T, KERN F, et al.miRCarta: A central repository for collecting miRNA candidates[J]. Nucleic Acids Research, 2018, 46(D1): D160-D167. DOI: 10.1093/nar/gkx851.

[16]AGARWAL V, BELL G W, NAM J W, et al. Predicting effective microRNA target sites in mammalian mRNAs[J]. elife, 2015, 4: e05005. DOI: 10.7554/eLife.05005.

[17]ENRIGHT A, JOHN B, GAUL U, et al. MicroRNA targets in drosophila[J]. Genome Biology, 2003, 4: 8. DOI: 10.1186/gb-2003-4-11-p8.

[18]STICHT C, DE LA TORRE C, PARVEEN A, et al. miRWalk: An online resource for prediction of microRNA binding sites[J]. PloS One, 2018,13(10): e0206239. DOI: 10.1371/journal.pone.0206239.

[19]HAUSCHILD A C, PASTRELLO C, EKAPUTERI G K A, et al. MirDIP 5.2: Tissue context annotation and novel microRNA curation[J]. Nucleic Acids Research, 2023, 51(D1): D217-D225. DOI: 10.1093/nar/gkac1070.

[20]LI Junhao, LIU Shun, ZHOU Hui, et al. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data[J]. Nucleic Acids Research, 2014, 42(D1): D92-D97. DOI: 10.1093/nar/gkt1248.

[21]CHEN Yuhao, WANG Xiaowei. miRDB: An online database for prediction of functional microRNA targets[J]. Nucleic Acids Research, 2020, 48(D1): D127-D131. DOI: 10.1093/nar/gkz757.

[22]WANG Xiaowei, EL NAQA I M. Prediction of both conserved and nonconserved microRNA targets in animals[J]. Bioinformatics, 2008, 24(3): 325-332. DOI: 10.1093/bioinformatics/btm595.

[23]LINSLEY P S, SCHELTER J, BURCHARD J, et al. Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression[J]. Molecular and Cellular Biology, 2007, 27(6): 2240-2252. DOI: 10.1128/MCB.02005-06.

[24]BARRETT T, TROUP D B, WILHITE S E, et al. NCBI GEO: Mining tens of millions of expression profiles—database and tools update[J]. Nucleic Acids Research, 2007, 35(suppl_1): D760-D765. DOI: 10.1093/nar/gkl887.

[25]WANG Xiaowei. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies[J]. Bioinformatics, 2016, 32(9): 1316-1322. DOI: 10.1093/bioinformatics/btw002.

[26]GROSSWENDT S, FILIPCHYK A, MANZANO M, et al. Unambiguous identification of miRNA: Target site interactions by different types of ligation reactions[J]. Molecular Cell, 2014, 54(6): 1042-1054. DOI: 10.1016/j.molcel.2014.03.049.

[27]HELWAK A, KUDLA G, DUDNAKOVA T, et al. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding[J]. Cell,2013,153(3):654-665. DOI: 10.1016/j.cell.2013.03.043.

[28]LIU Weijun, WANG Xiaowei. Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data[J]. Genome Biology, 2019, 20: 18. DOI: 10.1186/s13059-019-1629-z.

[29] MITRA R, BANDYOPADHYAY S. Multimitar: A novel multi objective optimization based mirnatarget prediction method[J]. PloS One, 2011, 6(9): e24583. DOI: 10.1371/journal.pone.0024583.

[30] KORFIATI A, THEOFILATOS K, KLEFTOGIANNIS D, et al. Predicting human mirna target genes using a novel computational intelligent framework[J]. Information Sciences, 2015, 294: 576-585. DOI: 10.1016/j.ins.2014.09.016.

[31] BANDYOPADHYAY S, MITRA R. TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples[J]. Bioinformatics, 2009, 25(20): 2625-2631. DOI: 10.1093/bioinformatics/btp503.

[32] XIAO Feifei, ZUO Zhixiang, CAI Guoshuai, et al. miRecords: an integrated resource for microRNA-target interactions[J]. Nucleic Acids Research, 2009, 37(suppl_1): D105-D110. DOI: 10.1093/nar/gkn851.

[33] BANDYOPADHYAY S, GHOSH D, MITRA R, et al. MBSTAR: Multiple instance learning for predicting specific functional binding sites in microRNA targets[J]. Scientific Reports, 2015, 5: 8004. DOI: 10.1038/srep08004.

[34] DING Jun, LI Xiaoman, HU Haijun. TarPmiR: A new approach for microRNA target site prediction[J]. Bioinformatics, 2016, 32(18): 2768-2775. DOI: 10.1093/bioinformatics/btw318.

[35] WONG L, YOU Zhuhong, GUO Zhenhao, et al. MIPDH: A novel computational model for predicting microRNA-mRNA interactions by DeepWalk on a heterogeneous network[J]. ACS Omega, 2020, 5(28): 17022-17032. DOI: 10.1021/acsomega.9b04195.

[36] KISHORE S, JASKIEWICZ L, BURGER L, et al. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins[J]. Nature Methods, 2011, 8(7): 559-564. DOI: 10.1038/NMETH.1608.

[37] PIÑERO J, BRAVO À, QUERALT-ROSINACH N, et al. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants[J]. Nucleic Acids Research, 2017,45(D1): D833-839. DOI: 10.1093/nar/gkw943.

[38] ZHOU Shijia, SUN Weicheng, ZHANG Ping, et al. Predicting pseudogene-miRNA associations based on feature fusion and graph auto-encoder[J]. Frontiers in Genetics, 2021, 12: 781277. DOI:10.3389/fgene.2021.781277.

[39] LI JunHao, LIU Shun, ZHOU Hui, et al. starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data[J]. Nucleic Acids Research, 2014, 42(D1): D92-D97. DOI: 10.1093/nar/gkt1248.

[40] FAN Chunyan, DING Mingchao. Inferring pseudogene-MiRNA associations based on an ensemble learning framework with similarity kernel fusion[J]. Scientific Reports, 2023,13:8833. DOI: 10.1038/s41598-023-36054-y.

[41] WANG Xinfei, YU Changqing, YOU Zhuhong, et al. A feature extraction method based on noise reduction for circRNA-miRNA interaction prediction combining multi-structure features in the association networks[J]. Briefings in Bioinformatics, 2023, 24(3): bbad111. DOI: 10.1093/bib/bbad111.

[42] LIU Ming, WANG Qian, SHEN Jian, et al. Circbank: A comprehensive database for circRNA with standard nomenclature[J]. RNA Biology, 2019, 16(7): 899-905. DOI: 10.1080/15476286.2019.1600395.

[43] WANG Xinfei, YU Changqing, LI Liping, et al. KGDCMI: A new approach for predicting circRNA-miRNA interactions from multi-source information extraction and deep learning[J]. Frontiers in Genetics, 2022, 13: 958096. DOI: 10.3389/fgene.2022.958096.

[44] ZHAO Zhengyang, LIN Jie, WANG Zhen, et al. SEBGLMA: Semantic embedded bipartite graph network for predicting lncRNA-miRNA associations[J]. International Journal of Intelligent Systems, 2023, 2023: 2785436. DOI: 10.1155/2023/2785436.

[45] GONG Jing, LIU Wei, ZHANG Jiayou, et al. lncRNASNP: A database of SNPs in lncRNAs and their potential functions in human and mouse[J]. Nucleic Acids Research, 2015, 43(D1):D181-D186. DOI:10.1093/nar/gku1000.

[46] PLA A, ZHONG Xiangfu, RAYNER S. miRAW: A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts[J]. PLoS Computational Biology, 2018, 14(7): e1006185. DOI: 10.1371/journal.pcbi.1006185.

[47] GU Tongjun, ZHAO Xiwu, BARBAZUK W B, et al. miTAR: A hybrid deep learning-based approach for predicting miRNA targets[J]. BMC Bioinformatics, 2021, 22: 96. DOI: 10.1186/s12859-021-04026-6.

[48] WEN Ming, CONG Peisheng, ZHANG Zhimin, et al. DeepMirTar: A deep-learning approach for predicting human miRNA targets[J]. Bioinformatics, 2018, 34(22): 3781-3787. DOI: 10.1093/bioinformatics/bty424.

[49] MIN S, LEE B, YOON S. TargetNet: Functional microRNA target prediction with deep neural networks[J]. Bioinformatics, 2022, 38(3): 671-677. DOI: 10.1093/bioinformatics/btab733.

[50] XIE Weidun, ZHENG Zetian, ZHANG Weitong, et al. SRG-vote: Predicting miRNA-gene relationships via embedding and LSTM ensemble[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(8): 4335-4344. DOI: 10.1109/JBHI.2022.3169542.

[51] XIE Weidun, LUO Jiawei, PAN Chu, et al. SG-LSTMFRAME: A computational frame using sequence and geometrical information via LSTM to predict miRNA-gene associations[J]. Briefings in Bioinformatics, 2021, 22(2): 2032-2042. DOI: 10.1093/bib/bbaa022.

[52] OUGHTRED R, RUST J, CHANG C, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions[J]. Protein Science, 2021, 30(1): 187-200. DOI:10.1002/pro.3978.

[53]SUN Jie, ZHOU Meng, YANG Haixun, et al. Inferring potential microRNA-microRNA associations based on targeting propensity and connectivity in the context of protein interaction network[J]. PloS One, 2013, 8(7): e69719. DOI: 10.1371/journal.pone.0069719.

[54]HWANG S, KIM C Y, YANG S, et al. HumanNet v2: Human gene networks for disease research[J]. Nucleic Acids Research, 2019, 47(D1): D573-D580. DOI: 10.1093/nar/gky1126.

[55]AHMED N K, ROSSI R A, LEE J B, et al. Role-based network embeddings[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(5): 2401-2415. DOI: 10.1109/TKDE.2020.3006475.

[56]LUO Jiawei, OUYANG Wenjue, SHEN Cong, et al. Multi-relation graph embedding for predicting mirna-target gene interactions by integrating gene sequence information[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(8): 4345-4353. DOI: 10.1109/JBHI.2022.3168008.

[57]SUN Yuzhuo, XIONG Fei, SUN Yongke, et al. A miRNA target prediction model based on distributed representation learning and deep learning[J]. Computational and Mathematical Methods in Medicine, 2022, 2022: 4490154. DOI: 10.1155/2022/4490154.

[58]GROSSWENDT S, FILIPCHYK A, MANZANO M, et al. Unambiguous identification of miRNA: Target site interactions by different types of ligation reactions[J]. Molecular Cell, 2014, 54(6): 1042-1054. DOI: 10.1016/j.molcel.2014.03.049.

[59]TALUKDER A, ZHANG Wencai, LI Xiaoman, et al. A deep learning method for miRNA/isomiR target detection[J]. Scientific Reports, 2022, 12: 10618. DOI: 10.1038/s41598-022-14890-8.

[60]YANG Tingpeng, WANG Yu, HE Yonghong. TEC-miTarget: enhancing microRNA target prediction based on deep learning of ribonucleic acid sequences[J]. BMC Bioinformatics, 2024, 25: 159. DOI: 10.1186/s12859-024-05780-z.

[61]LEE B. Deep learning-based microrna target prediction using experimental negative data[J]. IEEE Access, 2020, 8: 197908-197916. DOI: 10.1109/ACCESS.2020.3034681.

[62]LIU Pei, LIU Ying, LUO Jiawei, et al. MiRGraph: A transformer-based feature learning approach to identify microRNA-target interactions by integrating heterogeneous graph network and sequence information[EB/OL]. [2024-10-02]. https://doi.org/10.1101/2023.11.04.565620. DOI: 10.1101/2023.11.04.565620.

[63]LIU Pei, LUO Jiawei, CHEN Xiangtao. miRCom: tensor completion integrating multi-view information to deduce the potential disease-related miRNA-miRNA pairs[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2020, 19(3): 1747-1759. DOI: 10.1109/TCBB.2020.3037331.

[64]SZKLARCZYK D, GABLE A L, LYON D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets[J]. Nucleic ACIDS RESEARCH, 2019, 47(D1): D607-D613. DOI: 10.1093/nar/gky1131.

[65]HU Ziniu, DONG Yuxiao, WANG Kuansan, et al. Heterogeneous graph transformer[C]//WWW'20: Proceedings of The Web Conference 2020. New York, NY, United States: Association for Computing Machinery, 2020: 2704-2710. DOI: 1145/3366423.3380027.

[66]ZHANG Jialin, ZHU Haoran, LIU Yin, et al. miTDS: Uncovering miRNA-mRNA interactions with deep learning for functional target prediction[J]. Methods, 2024, 223: 65-74. DOI: 10.1016/j.ymeth.2024.01.011.

[67]COCK P J A, ANTAO T, CHANG J T, et al. Biopython: Freely available Python tools for computational molecular biology and bioinformatics[J]. Bioinformatics, 2009, 25(11): 1422. DOI: 10.1093/bioinformatics/btp163.

[68]SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 1-9. DOI: 10.1109/cvpr.2015.7298594.

[69]JIANG Hui, WANG Jiaxin, LI Min, et al. miRTRS: A recommendation algorithm for predicting miRNA targets[J]. IEEE/ACM Transactions on Computational Biology And Bioinformatics, 2020, 17(3): 1032-1041. DOI: 10.1109/TCBB.2018.2873299.

[70]JIANG Hui, YANG Mengyun, CHEN Xiang, et al. miRT-MC: A miRNA target prediction method based on matrix completion algorithm[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 24(12): 3630-3641. DOI: 10.1109/JBHI.2020.2987034.

[71]TALUKDER A, LI Xiaoman, HU Haiyan. Position-wise binding preference is important for miRNA target site prediction[J]. Bioinformatics, 2020, 36(12): 3680-3686. DOI: 10.1093/bioinformatics/btaa195.