

# 基于深度学习的长链非编码RNA 与微小RNA相互作用预测的 研究进展

王文雅

辽宁科技大学计算机与软件工程学院, 辽宁 鞍山

收稿日期: 2022年3月6日; 录用日期: 2022年4月6日; 发布日期: 2022年4月15日

## 摘要

长链非编码RNA (lncRNA)与微小RNA (miRNA)都是非编码RNA,越来越多的证据表明两者之间的相互作用与癌症的发展、基因调控、细胞代谢等生物学过程高度相关。与此同时,随着RNA序列技术的快速发展人们发现了许多新的lncRNA和miRNA,这可能有助于探索新的基因调控模式,人们对于lncRNA-miRNA相互作用的研究兴趣也随之越来越大。为此,我们回顾了目前lncRNA-miRNA相互作用关系预测的研究进展,我们针对部分研究人员的最新研究成果着重调查了他们使用的计算方法和数据库。调查结果显示深度学习已经成为lncRNA-miRNA相互作用关系预测的首选策略,这可能是由于深度学习基础设施和专业知识的快速增长。虽然这些方法中有许多都有明显的局限性,但深度学习有望在未来lncRNA-miRNA相互作用关系预测的领域取得更加充分的应用。

## 关键词

lncRNA-miRNA相互作用, 深度学习, 数据库, 计算框架

## Research Progress of Predicting Long Non-Coding RNA-MicroRNA Interaction Based on Deep Learning

Wenya Wang

School of Computer Science and Software Engineering, University of Science and Technology Liaoning,  
Anshan Liaoning

Received: Mar. 6<sup>th</sup>, 2022; accepted: Apr. 6<sup>th</sup>, 2022; published: Apr. 15<sup>th</sup>, 2022

文章引用: 王文雅. 基于深度学习的长链非编码RNA与微小RNA相互作用预测的研究进展[J]. 数据挖掘, 2022, 12(2): 152-160. DOI: 10.12677/hjdm.2022.122016

## Abstract

Long non-coding RNA (lncRNA) and microRNA (miRNA) are both noncoding RNA. More and more evidence shows that the interaction between them is highly related to biological processes such as cancer development, gene regulation and cell metabolism. At the same time, with the rapid development of RNA sequence measuring technology, many new lncRNAs and miRNAs have been found, which may help to explore new gene regulation modes, and people are more and more interested in the research of lncRNA-miRNA interaction. Therefore, we reviewed the current research progress in the prediction of lncRNA-miRNA interaction. We focused on the calculation methods and databases used by some researchers according to their latest research results. The results show that deep learning has become the preferred strategy for the prediction of lncRNA-miRNA interaction, which may be due to the rapid growth of deep learning infrastructure and expertise. Although many of these methods have obvious limitations, deep learning is expected to become the basis of modern lncRNA-miRNA interaction prediction algorithms.

## Keywords

lncRNA-miRNA Interactions, Deep Learning, Database, Computational Frame

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 介绍

非编码 RNA (ncRNAs) [1] 虽然不能编码蛋白质, 但是它们在许多生命过程中发挥着不可或缺的作用 [2]-[7]。研究表明, 许多 ncRNA 参与了各种生命调节过程 [8] [9], lncRNA 和 miRNA 作为两种典型的 ncRNA 被证明与癌症的发展、基因调控、细胞代谢等生物学过程有关。lncRNA 是一种通常超过 200nts 的非编码转录本, 其具有非常重要的调控功能且几乎参与到了各种生物学过程, 与各种疾病的发生发展紧密关联, 从而成为过去几年和将来的研究热点和重点。对于人类基因组来说, 产生的 lncRNA 数量比编码 RNA 的数量要多得多, 目前除了少数 lncRNA 的功能比较明确外, 大部分 lncRNA 的功能都还未知。它们作为 miRNA 的靶模拟物可以对 miRNA 进行调节, miRNA 是一类长度约为 20nts 的非编码 RNA 物种, 在蛋白质编码基因的转录后调控中发挥核心作用。因此, 探究 lncRNA-miRNA 的作用关系对研究 lncRNA 功能来说非常必要。目前, 已经可以利用大规模实验方法验证 lncRNA-miRNA 潜在关联, 但这些实验往往会浪费大量的时间成本和物质成本。鉴于这些情况, 研究人员已经致力于将机器学习以及深度学习利用到两者的关系预测的工作中, 并取得了较为不错的结果。在这篇综述中, 我们特别考虑了 lncRNA 与 miRNA 的相互作用, 总结了近期关于两者相互作用关系预测的科研文献, 并做了简要的结果对比来直观展示目前部分预测模型的性能。借此帮助读者更加充分地了解该领域的科研进展。

## 2. 数据集

MiRBase [10]、CANTATAdb、lncRNASNP [11]、GENCODE [12]、GreeNC、StarBase 是一些常见的数据库, 这些数据库包含 miRNA 及 lncRNA 的序列信息与 lncRNA-miRNA 相互作用相关的信息, 这些数据都是通过实验室分析、计算分析和文献挖掘获得的, 具有较高可信度。目前有两大类: 包含各类

物种的 miRNA、lncRNA 的序列信息的数据库和包含 lncRNA 和 miRNA 相互作用关系的数据库。

Starbase 是由中山大学开发的中国生物信息学最具影响力的数据库, 该数据库用过实验获得了丰富的 miRNA-ncRNA、miRNA-mRNA、RBP-RNA 和 RNA-RNA 的数据。LncRNASNP 是由华中科技大学郭安源教授团队所开发, 提供不同类型的相关数据资源, 包括 lncRNA 表达谱、扩展的 lncRNA 相关疾病和 lncRNA 的非编码变异。Starbase、LncRNASNP 都包含了 lncRNA-miRNA 相互作用的数据, 以及与相互作用相关的许多附加属性, 且并不仅仅局限于 lncRNA, 还包含各种其他的相互作用信息。

MiRBase、GENCODE、GreeNC、CANTATAdb 包含 lncRNA 与 miRNA 序列的详细信息。虽然不直接包含作用关系信息, 但这些可以为特征提取工作提供基础数据从而进一步提高预测性能, 是 lncRNA-miRNA 相互作用关系预测研究中必不可少的基础数据。

### 3. 评价指标

在使用深度学习算法过程中, 针对不同的问题需要采用不同的模型评估标准, 主要有分类和回归两大类。分类问题采用混淆矩阵、准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1 score、ROC 曲线来作为模型评估指标。混淆矩阵是监督学习中的一种可视化工具, 主要用于比较分类结果和实例的真实信息, 矩阵中的每一行代表实例的预测类别, 每一列代表实例的真实类别。TP 代表实际为正类且预测为正类的样本, FP 代表实际为负类而预测为正类的样本, TN 代表实际为负类且预测也为负类的样本, FN 代表实际为正类而预测为负类的样本。混淆矩阵如图 1 所示:

		actual value	
		TP (True Positive)	FP (False Positive)
prediction outcome	FN (False Negative)	TN (True Negative)	

Figure 1. Confusion Matrix

图 1. 混淆矩阵

准确率(Accuracy)是指被预测分类正确的样本占总样本的比; 精确率(Precision)是指被预测为正样本的数据中, 实际也为正样本的占比; 召回率(Recall)是指在实际正样本中被预测正确的样本占比; F-值(F1 score)是 Precision 和 Recall 的调和平均值。以上评价指标的公式如下:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} = 2 \cdot \frac{PRE \cdot REC}{PRE + REC}$$

其中 ROC 曲线为受试者工作特性曲线,横坐标为负正类率(FPR),纵坐标为真正类率(TPR),AUC 为 ROC 曲线下的面积,取值在(0, 1)之间,且 AUC 值越接近 1,ROC 曲线越接近左上角,该分类器的性能越好。FPR 与 TPR 的公式如下:

$$FPR = \frac{TP}{TP + FN}$$

$$TPR = \frac{FP}{FP + TN}$$

#### 4. 预测算法

大多数 lncRNA-miRNA 相互作用关系预测算法需要利用由相关数据库获取到的数据来进行预测。lncRNA-miRNA 相互作用关系预测的计算策略主要分为机器学习和深度学习两类,预测方法主要包括深度学习、基于图的方法、相似性网络、基于树的方法。以往的机器学习是利用 lncRNA 和 miRNA 分子序列信息之间已知的相互作用,利用已知的二级结构来改善它们的性能表达。在过去的几年中,出现了大量的基于深度学习的预测算法,深度学习提出了一种让计算机自动学习出数据特征的方法,并将特征学习融入到了建立模型的过程中,从而减少了人为设计特征造成的不完备性,但是为了能使数据的特征被无偏差学习,我们需要为深度学习的模型提供大量数据,以保证模型性能的提高。

Huang 等人在 2019 年提出了一种通过结合图卷积神经网络(GNN)和自动编码器技术(AutoEncoder)的端到端预测算法,称为 GCLMI [13]。该方法不需要对特征信息进行任何预处理,就可以将节点属性的原始数据与交互网络的拓扑结构结合起来。基于从公共数据库中收集到的真实数据集,多倍交叉验证的实验结果表明了该预测模型预测性能的鲁棒性和有效性。该方法证明了该模型中所设计的图卷积层能够通过对具有节点特征的图进行过滤来有效地整合输入数据。

在该算法中将 lncRNA-miRNA 相互作用的预测任务转换为异构二部图上的链接预测问题。构建一个相邻矩阵其形状是  $N_l \times N_m$  的图 M,其中  $N_l$ 、 $N_m$  分别为 lncRNA 及 miRNA 节点数。该图中的  $M_{ij}$  为第 i 类 lncRNA 和第 j 类 miRNA 之间是否具有相互作用,该数据由公共数据库获取。预测任务可以看作是利用深度学习来获取已有的边的规律从而来预测判断图中未知的边是否存在。GCLMI 的整体流程如图 2 所示:

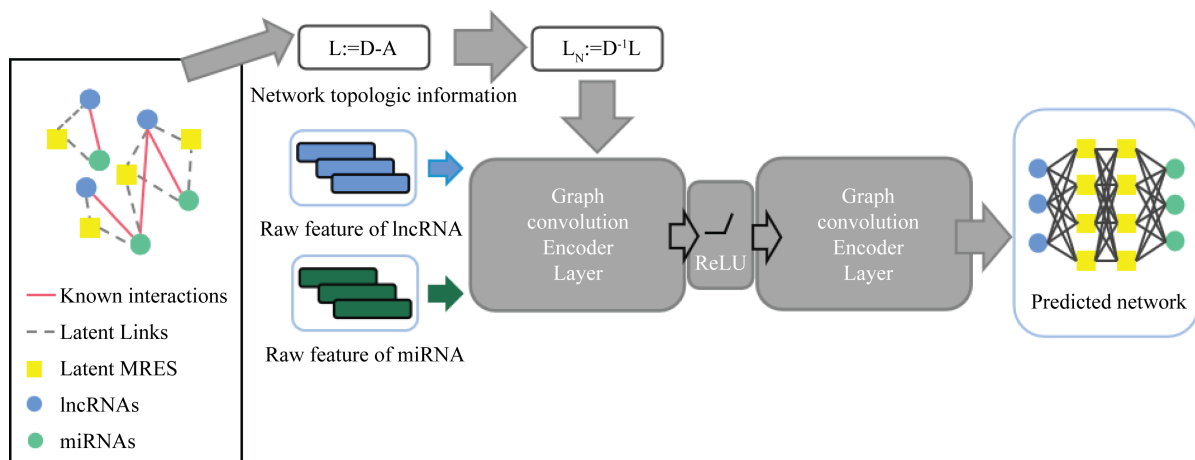


Figure 2. The overall flow chart of GCLMI

图 2. GCLMI 的整体流程图

Kang 等人在 2020 年提出了一种基于混合模型和模糊决策来实现植物 lncRNA-miRNA 相互作用关系预测的方法, 称为 PmliPred [14]。该方法由卷积神经网络(CNN)和双向门控递归单元(BiGRU)集成构建了一个混合模型。CNN 是一个采用多层次结构且具有鲁棒性的深度学习方法, 可以从二维数据中提取到拓扑结构, 采用反向传播机制来优化整个网络结构, 从而自动求解网络中的未知参数。BiGRU 是一个循环神经网络为 LSTM 的变种, 可以解决长期记忆和反向传播中的梯度问题。

该方法采取了 one-hot 编码来获取原始序列信息中分子之间的关系, 编码后得到的二维矩阵可作为模型输入, CNN 可以提取抽象特征, 从不同层次的原始序列中通过卷积操作和压缩后平铺成一维向量输入到 GRU 单元, 模糊决策解决了模型在做决策时的不确定性和不精确性。最后对 PmliPred 的参数和权重策略进行调整, 以使性能最大化。PmliPred 的整体流程如图 3 所示:

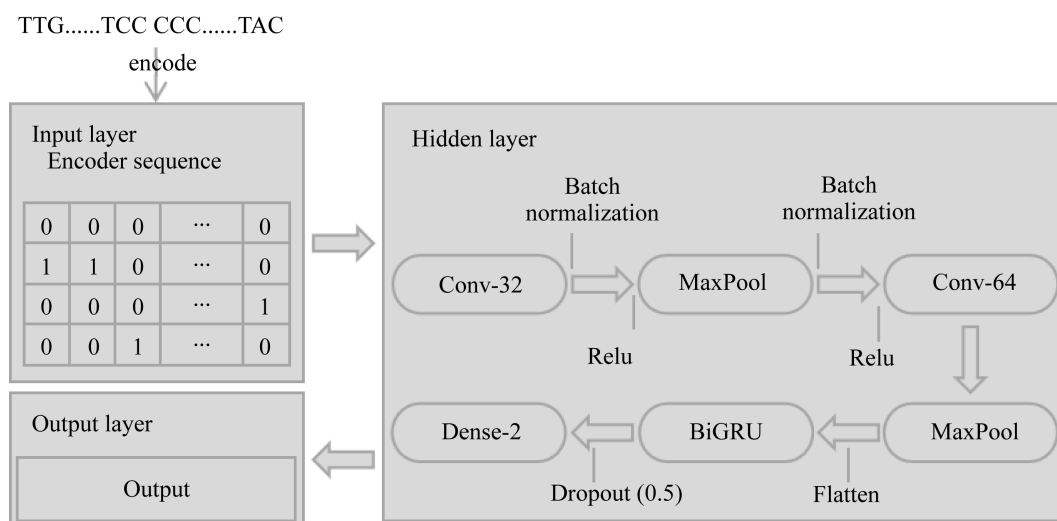


Figure 3. The overall flow chart of PmliPred  
图 3. PmliPred 的整体流程图

Wang 等人在 2020 年提出了一个基于混合序列特征及 CNN 的混合模型, 称为 LncMirNet [15]。该模型引入 k-mer [16]、CTD [17]、doc2vec [18]和 role2vec [19]四种基于序列的特征来编码 lncRNA/miRNA 序列, 然后采用直方图方法将多种类型的特征融合到一个矩阵, 最后将矩阵输入 CNN 以达到对 lncRNA-miRNA 相互作用关系的预测。

首先基于 lncRNA/miRNA 的核苷酸序列信息, 采用 k-mer 来表示一个基本但却不可或缺的特征, 计算了 lncRNA 的四种特征, 分别为 1-mer、2-mer、3-mer、4-mer, 由于 miRNA 序列过短其 4-mer 特征通常很稀疏因此只计算 miRNA 的 1-mer、2-mer、3-mer 等特征。最后, 将四种 k-mer 特征合并为一个具有 340 维度的向量。采用 CTD 特征来表示 RNA 的结构信息, 具有 30 维的 CTD 特征来自组成、过渡和分布, 其中组成特性是指特定特性的氨基酸的数量除以氨基酸的总数, 过渡特征描述特定特性的氨基酸的百分比频率, 而分布特征测量的是一个特定属性的第一个、25%、50%、75%和 100%的氨基酸所在的链长。采用 Doc2vec 构建 RNA 序列的分布式表示特征, 使用局部上下文和句子全局信息来学习句子表示, 任何的 RNA 序列都编码到一个固定大小的载体中。采用 Role2vec 结合作用对数据集对每个节点进行嵌入, 因此 Role2vec 结合了图结构和节点属性信息, 以学习每个节点的表示。最后依次将 lncRNAs/miRNAs 的 k-mer、CTD、doc2ve 和图嵌入特征融合到一个矩阵中, 通过直方图模拟 CNN 学习模式将多个特征向量融合成一个直方图矩阵。最后, 将直方图矩阵输入到 CNN 模型中, CNN 应用卷积核从原始输入数据

矩阵中自动提取潜在的特征，是一个强大的解决分类和监管问题的算法。LncMirNet 的整体工作流程如图 4 所示：

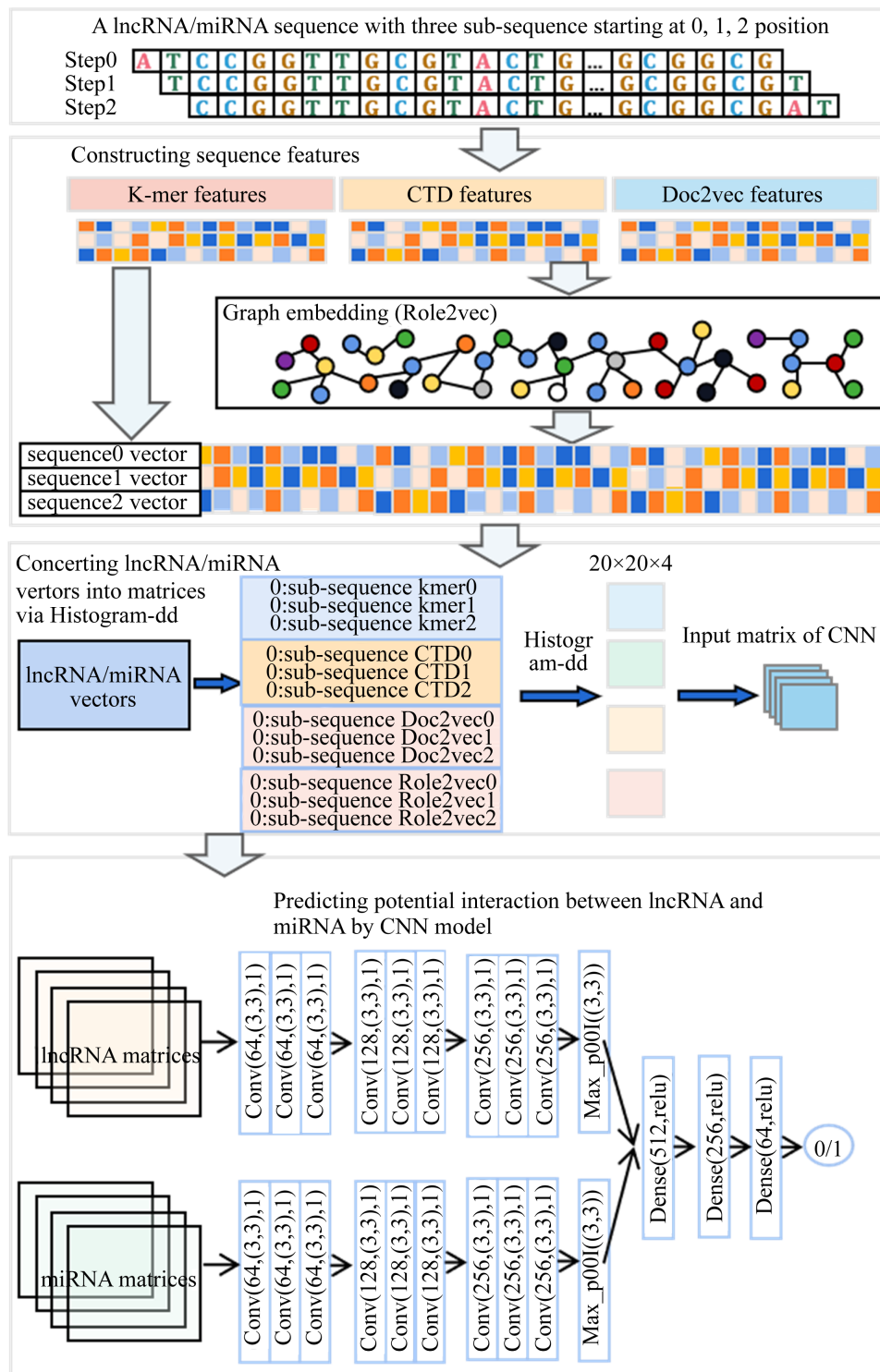


Figure 4. The overall flow chart of LncMirNet

图 4. LncMirNet 的整体流程图



Yu 等人在 2021 年提出了一种基于 rna2vec 的预训练和深度特征挖掘机制的深度学习模型，称为 preMLI [20]。该模型的三个主要步骤分别是预训练、深度特征挖掘和相关预测。

首先基于 Word2vec 作了部分调整提出了一个新的词训练方法 rna2vec，将 miRNA 和 lncRNA 数据集分别作为 rna2vec 的输入进行预训练，然后利用得到的权值矩阵将 RNA 的表达嵌入到训练数据集中。然后由于在自然语言处理的相关研究中，研究者经常使用递归神经网络(RNNs)来分析序列，与 CNN 相比 RNN 模型可以解决长期依赖问题。但在这个问题中，因为 CNN 可以从短序列片段中获得更多的潜在特征并具有较强的并行计算能力，所以该模型模型使用了包括 CNN 和双向门控循环单元(Bi-GRU)集成的混合模型。其中，利用 CNN 学习 lncRNA 和 miRNA 的局部特征，BiGRU 捕获局部特征的长期依赖性。在网络的下游，该模型还增加了一个注意力机制来关注关键特征。

由于 miRNA 和 lncRNA 序列长度的差异，在序列嵌入后直接剪接特征向量并将其输入下游任务这样可能会导致 miRNA 的潜在特征获取不足。对此模型设计了一种深度特征挖掘机制，即首先通过上述网络从嵌入的输出特征向量中获得足够的潜在特征，最后将该向量连接起来作为预测层的输入。并且在两个网络上基于多个测试集进行了实验。实验表明，预训练机制和深度特征挖掘机制对模型的预测性能都有积极的影响，预训练可以提供更准确的单词向量表示，深度特征挖掘机制也提高了模型的预测性能。preMLI 的整体工作流程如图 5 所示：

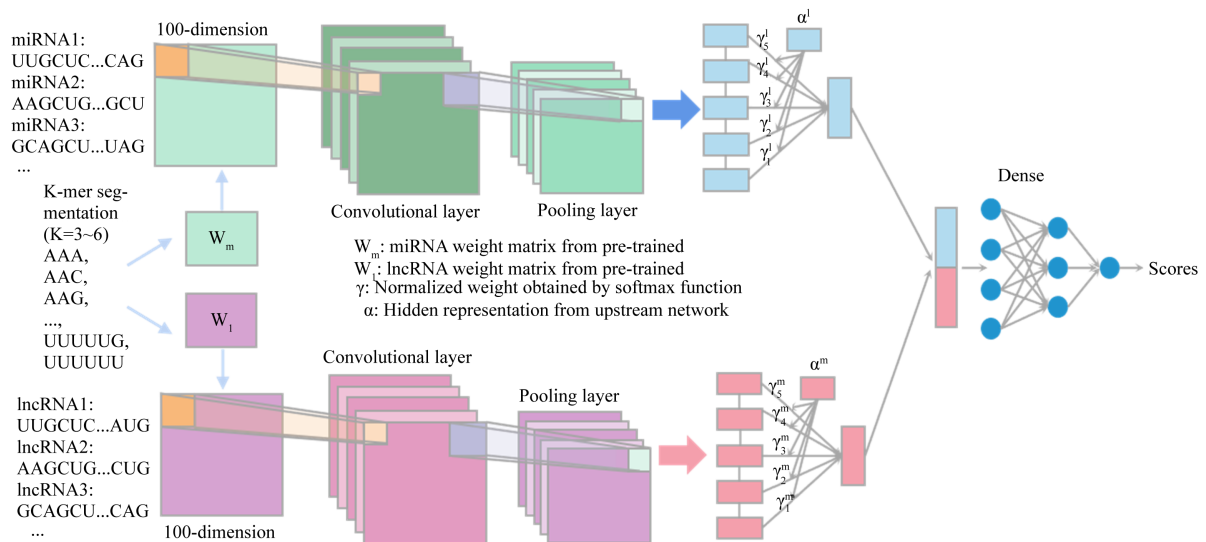


Figure 5. The overall flow chart of preMLI

图 5. preMLI 的整体流程图

### 5. 结果对比

我们搜集到以上介绍的模型所提供的数据集及代码，分别对其进行复现得到对应的评价指标 AUC 的值。并且将以上介绍的四种方法所使用到的模型以及结果进行对比，四种模型都取得了较为不错的结果，可以说明这四种模型都能够较为有效的预测出潜在的 lncRNA-miRNA 作用关系。但在对比中可以看出模型 LncMirNet 及 preMLI 的结果比余下两者模型的结果要高出许多，均在 0.9 以上。综合上述对模型的详细介绍中我们可以发现 LncMirNet 及 preMLI 在数据处理阶段都采用了较为独特的编码方式，尽可能的利用编码方式获取到更加完备的 lncRNA 与 miRNA 的序列信息，由此可见信息的特征表示对于预测结果的影响很大。四种深度学习模型在 lncRNA 与 miRNA 作用关系预测表现上的比较如表 1 所示：

**Table 1.** Standard experimental system result data  
**表 1.** 标准实验系统结果数据

Method	Classifier	Database	Performance (AUC)
GCLMI	GNN、AutoEncoder	lncRNASNP、starBase	0.8567 ± 0.0009
PmliPred	CNN、BiGRU	CANTATAdb2.0、miRBase、GreeNC	0.8386
LncMirNet	Role2vec、CNN	lncRNASNP2、miRbase、GENCODE	0.9381
preMLI	rna2vec、CNN、BiGRU	CANTATAdb、miRBase、GreeNC	0.9773

## 6. 结论与展望

实验研究证明 lncRNA 与 miRNA 相互作用与癌症的发展、基因调控、细胞代谢等生物学过程高度相关, 因此探究 lncRNA 与 miRNA 之间的相互作用关系是十分有必要的。在早期研究中, 研究者主要通过实验室实验探索未知的 lncRNA-miRNA 相互作用关系, 然而由于生物实验室发现 lncRNA 和 miRNA 之间的潜在相互作用的工作是劳动密集型、耗时且昂贵的, 所以有越来越多的计算方法被用来辅助探索 lncRNA 和 miRNA 之间的潜在作用关系。但是由于受到生物学实验的限制, 现有的 lncRNA 与 miRNA 关联数据十分有限, 这使得对于深度学习模型来说样本数量小, 从而导致计算结果性能低。另一方面由于我们现下只拥有正样本数据集对于负样本的选取采用的是随机方式, 则是除去正样本之后随机选取需要数量的负样本, 这导致负样本数据中可能掺杂着未被证实的正样本, 从而使得模型结果存在一定误差。对于这些问题我们提出了相应的解决方式: 1) 选择规模更大数据量更加丰富的数据集, 提高样本量进而提高模型性能; 2) 将负样本中被初步预测为正样本的数据进行剔除, 尽可能保留真正的负样本从而减少实验误差; 3) 寻找更加适合自己数据结构的深度学习模型。

在这篇综述中, 我们回顾了近几年来使用深度学习工具来进行 lncRNA-miRNA 相互作用关系预测的一些方法, 许多计算方法已经成功地充分应用于辅助生物实验, 并广泛应用于生物信息学。此外我们也介绍了目前预测研究存在的问题, 并为未来从事此研究方向的学者提出了一些建议, 希望这些能帮助改进现有的方法来取得更佳的结果。

## 参考文献

- [1] Yang, S., Wang, Y., Zhang, S., Hu, X., Ma, Q. and Tian, Y. (2020) NCResNet: Noncoding Ribonucleic Acid PredictionBased on a Deep Resident Network of Ribonucleic Acid Sequences. *Frontiers in Genetics*, **11**, Article No. 90. <https://doi.org/10.3389/fgene.2020.00090>
- [2] Huang, J.Z., Chen, M., Chen, D., Gao, X.C., Zhu, S., Huang, H., Hu, M., Zhu, H. and Yan, G.R. (2017) A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Molecular Cell*, **68**, 171-184.e6. <https://doi.org/10.1016/j.molcel.2017.09.015>
- [3] Chakravarty, D., Sboner, A., Nair, S.S., Giannopoulou, E., Li, R., Hennig, S., Mosquera, J.M., Pauwels, J., Park, K., Kossai, M., et al. (2014) The Oestrogen Receptor Alpha-Regulated lncRNA NEAT1 Is a Critical Modulator of Prostate Cancer. *Nature Communications*, **5**, Article No. 5383. <https://doi.org/10.1038/ncomms6383>
- [4] Bhan, A. and Mandal, S.S. (2015) lncRNA HOTAIR: A Master Regulator of Chromatin Dynamics and Cancer. *Biochimica et Biophysica Acta (BBA)—Reviews on Cancer*, **1856**, 151-164. <https://doi.org/10.1016/j.bbcan.2015.07.001>
- [5] Sun, M., Nie, F., Wang, Y., Zhang, Z., Hou, J., He, D., Xie, M., Xu, L., De, W., Wang, Z., et al. (2016) lncRNA-HOXA11-AS Promotes Proliferation and Invasion of Gastric Cancer by Scaffolding the Chromatin Modification Factors PRC2, LSD1, and DNMT1. *Cancer Research*, **76**, 6299-6310. <https://doi.org/10.1158/0008-5472.CAN-16-0356>
- [6] Xue, X., Yang, Y.A., Zhang, A., Fong, K.W., Kim, J., Song, B., Li, S., Zhao, J.C. and Yu, J. (2016) lncRNA HOTAIRenhances ER Signaling and Confers Tamoxifen Resistance in Breast Cancer. *Oncogene*, **35**, 2746-2755. <https://doi.org/10.1038/onc.2015.340>
- [7] Cao, S., Liu, W., Li, F., Zhao, W. and Qin, C. (2014) Decreased Expression of lncRNA GAS5 Predicts a Poor Prognosis



- sis Incervical Cancer. *International Journal of Clinical and Experimental Pathology*, **7**, 6776-6783.
- [8] Cai, Y., Yu, X., Hu, S. and Yu, J. (2009) A Brief Review on the Mechanisms of miRNA Regulation. *Genomics, Proteomics & Bioinformatics*, **7**, 147-154. [https://doi.org/10.1016/S1672-0229\(08\)60044-3](https://doi.org/10.1016/S1672-0229(08)60044-3)
- [9] Peng, W.X., Koirala, P. and Mo, Y.Y. (2017) LncRNA-Mediated Regulation of Cell Signaling in Cancer. *Oncogene*, **36**, 5661-5667. <https://doi.org/10.1038/onc.2017.184>
- [10] Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019) GENCODE Reference Annotation for the Human and Mouse Genomes. *Nucleic Acids Research*, **47**, D766-D773. <https://doi.org/10.1093/nar/gky955>
- [11] Miao, Y.R., Liu, W., Zhang, Q. and Guo, A.Y. (2018) LncRNASNP2: An Updated Database of Functional SNPs and Mutations in Human and Mouse lncRNAs. *Nucleic Acids Research*, **46**, D276-D280. <https://doi.org/10.1093/nar/gkx1004>
- [12] Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) miRBase: From MicroRNA Sequences to Function. *Nucleic Acids Research*, **47**, D155-D162. <https://doi.org/10.1093/nar/gky1141>
- [13] Huang, Y.A., Huang, Z.A., You, Z.H., Zhu, Z., Huang, W.-Z., Guo, J.-X., et al. (2019) Predicting lncRNA-miRNA Interaction via Graph Convolution Auto-Encoder. *Frontiers in Genetics*, **10**, Article No. 758. <https://doi.org/10.3389/fgene.2019.00758>
- [14] Kang, Q., Meng, J., Cui, J., Luan, Y. and Chen, M. (2020) PmlPred: A Method Based on Hybrid Model and Fuzzy Decision for Plant miRNA-lncRNA Interaction Prediction. *Bioinformatics*, **36**, 2986-2992. <https://doi.org/10.1093/bioinformatics/btaa074>
- [15] Yang, S., Wang, Y., Lin, Y., Shao, D., He, K. and Huang, L. (2020) LncMirNet: Predicting lncRNA-miRNA Interaction Based on Deep Learning of Ribonucleic Acid Sequences. *Molecules*, **25**, Article No. 4372. <https://doi.org/10.3390/molecules25194372>
- [16] Tong, X. and Liu, S. (2019) CPPred: Coding Potential Prediction Based on the Global Description of RNA Sequence. *Nucleic Acids Research*, **47**, e43. <https://doi.org/10.1093/nar/gkz087>
- [17] Liu, B., Fang, L., Wang, S., Wang, X., Li, H. and Chou, K.C. (2015) Identification of MicroRNA Precursor with The-degenerate K-Tuple or Kmer Strategy. *Journal of Theoretical Biology*, **385**, 153-159. <https://doi.org/10.1016/j.jtbi.2015.08.025>
- [18] Lau, J.H. and Baldwin, T. (2016) An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation. *Proceedings of the 1st Workshop on Representation Learning for NLP*, Berlin, 11 August 2016, 78-86. <https://doi.org/10.18653/v1/W16-1609>
- [19] Ahmed, N., Rossi, R.A., Lee, J., Willke, T., Zhou, R., Kong, X. and Eldardiry, H. (2020) Role-Based Graph Embeddings. *IEEE Transactions on Knowledge and Data Engineering*, **34**, 2401-2415. <https://doi.org/10.1109/TKDE.2020.3006475>
- [20] Yu, X., Jiang, L., Jin, S., Zeng, X. and Liu, X. (2022) preMLI: A Pre-Trained Method to Uncover MicroRNA-lncRNA Potential Interactions. *Briefings in Bioinformatics*, **23**, Article No. bbab470. <https://doi.org/10.1093/bib/bbab470>