

基于深度学习融合模型的蛋白质-蛋白质互作预测研究

白跃花, 蔡景耀, 林佳*

华北理工大学生命科学院, 河北 唐山

收稿日期: 2025年4月10日; 录用日期: 2025年5月21日; 发布日期: 2025年5月30日

摘要

蛋白质间相互作用(protein-protein interaction, PPI)是维系细胞正常生理功能的核心要素,明确参与代谢调控、信号转导、免疫应答等诸多关键生物学进程,与人类健康紧密相关。深入探究PPI对于阐释疾病发生发展机制、挖掘潜在药物靶点意义重大。传统PPI预测方法多依赖实验手段,存在通量低、假阳性高以及难以处理复杂互作模式等问题。随着人工智能技术蓬勃发展,深度学习凭借强大特征学习与模态融合能力,逐渐成为PPI预测新锐手段。鉴于此,本研究创新性提出一种多模态融合的深度神经网络模型,有机整合蛋白质序列与结构特征,充分挖掘二者关联模式,实现对蛋白质间潜在互作关系精准预测。经多个权威基准数据集验证,该模型在准确率、召回率、AUC值等核心评估指标上全面超越现有方法。这一成果有望为蛋白质互作网络解析提供高效工具,为细胞信号传导规律探索、疾病分子机制研究注入新动力,进而为精准医学发展与药物研发进程加速助力。

关键词

蛋白质互作, 深度学习融合模型, PPI

Protein-Protein Interaction Prediction Based on a Deep Learning Fusion Model

Yuehua Bai, Jingyao Cai, Jia Lin*

College of Life Sciences, North China University of Science and Technology, Tangshan Hebei

Received: Apr. 10th, 2025; accepted: May 21st, 2025; published: May 30th, 2025

Abstract

Protein-protein interaction (PPI) is the core elements that maintain normal cellular physiological

*通讯作者。

文章引用: 白跃花, 蔡景耀, 林佳. 基于深度学习融合模型的蛋白质-蛋白质互作预测研究[J]. 生物医学, 2025, 15(3): 577-585. DOI: 10.12677/hjbm.2025.153066

functions, clearly participating in many critical biological processes such as metabolic regulation, signal transduction, and immune response, which are closely related to human health. In-depth investigation of PPI is crucial for elucidating disease mechanisms and uncovering potential drug targets. Traditional PPI prediction methods often rely on experimental approaches, facing issues like low throughput, high false positives, and difficulty in handling complex interaction patterns. With the rapid development of artificial intelligence technology, deep learning has emerged as a new and powerful tool for PPI prediction, thanks to its strong feature learning and modal fusion capabilities. In light of this, our study innovatively proposes a multi-modal fusion deep neural network model that organically integrates protein sequence and structural features, fully exploring their associated patterns to achieve precise prediction of potential protein interactions. Validation on multiple authoritative benchmark datasets shows that this model outperforms existing methods comprehensively in core evaluation metrics such as accuracy, recall rate, and AUC value. This achievement is expected to provide an efficient tool for protein interaction network analysis, inject new momentum into the exploration of cellular signaling rules and disease molecular mechanisms, and accelerate the development of precision medicine and drug research.

Keywords

Protein-Protein Interaction (PPI), Deep Learning Fusion Model, PPI Prediction

Copyright © 2025 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. 引言

蛋白质 - 蛋白质相互作用(PPI)构成了细胞生命活动的核心机制, 主要通过共价键的形成以及非共价表面接触来实现。这些相互作用方式不仅构成了蛋白质功能的基础, 还通过调控蛋白质复合体的组装与稳定性, 对细胞内的多种生物学过程产生影响[1]。PPI 在高等动物胚胎发育、肿瘤发生以及信号转导通路中扮演着关键角色, 信号转导通路作为细胞命运决定、组织稳态维持及病理状态转化的核心调控机制, 一直是生物医学领域的研究热点[2]。PPI 网络的失调与癌症、神经退行性疾病等复杂疾病的病理进程密切相关[3] [4]。因此, 精准预测 PPI 对于解析疾病机制、开发靶向药物具有重要科学价值, 并可助力疾病诊断和治疗研究。

2. 蛋白质 - 蛋白质互作概述

2.1. 蛋白质互作的功能意义

蛋白质并非孤立发挥作用, 其功能实现往往依赖于与其他生物分子(如蛋白质、核酸或配体)的动态结合, 形成功能复合体以执行特定任务。这种相互作用在真核生物(如人类、酵母)与原核生物中均具有高度保守性, 是细胞分裂、能量代谢及免疫防御等生命过程的核心调控基础。例如, 酶促反应中底物的催化依赖于酶与底物的特异性结合, 而信号通路的激活则通过受体蛋白与配体的互作触发下游级联反应[5]。

2.2. 研究目的

全基因组序列信息虽提供了丰富的遗传基础, 但仅凭其无法充分解释细胞的复杂生命活动现象。蛋白质作为细胞活性和功能的最终执行者, 其相互作用(PPI)是实现细胞功能的核心机制之一[6]。本研究旨在构建一种基于深度学习的 PPI 预测框架, 通过整合序列与结构特征突破传统方法的局限性。具体目标

包括：1) 研究一种基于深度学习的 PPI 预测模型；2) 通过整合蛋白质序列和结构信息，提高 PPI 预测的准确性和可靠性；3) 评估该模型在 PPI 预测中的性能，并与传统方法进行比较。

2.3. 蛋白质互作的机制

蛋白质之间的互作主要通过其表面的特定区域(即互作位点)来实现。这些结合位点由若干氨基酸残基构成，通过氢键、疏水效应及静电吸引等非共价作用维持复合体的稳定性。氢键是分子间或分子内由电负性原子(如 O、N、F)与相邻 H 原子形成的静电吸引作用，强度介于范德华力和化学键之间。疏水作用是蛋白质的关键特性之一，对维持其折叠稳定性和生物活性至关重要，同时也是基于疏水性差异的蛋白质分离技术的重要理论基础。而静电相互作用在蛋白质与多糖的结合中表现为两种形式：一种是通过共价键形成的稳定连接，另一种则是基于静电吸引力的非共价相互作用[7]。近年来，随着深度学习技术的发展，研究人员开始利用这些技术来预测 PPI，通过学习蛋白质序列和结构中的复杂模式来提高预测的准确性。

2.4. PPI 研究的技术瓶颈

传统实验手段(如酵母双杂交、免疫共沉淀)虽能验证特定互作，但存在通量低、假阳性高等缺陷[8]。而现有计算模型(如支持向量机)在复杂网络建模中难以捕捉动态互作特征。此外，蛋白质的三维结构信息在 PPI 预测中的应用仍有限，而三维结构是决定互作特异性的关键因素之一。总之，多模态数据的整合仍是一个挑战，如何有效融合蛋白质序列、结构、基因表达和代谢数据等信息是当前研究的重要课题。

3. 蛋白质结构特征与互作关系

3.1. 一级结构对于蛋白质结构的影响

一级结构是蛋白质互作的基础，其氨基酸序列决定了蛋白质的折叠方式和功能位点。关键残基直接参与结合，如酶活性中心的催化氨基酸或信号肽中的特异序列(如 RGD 基序)，能够被特定受体识别。序列变异可能破坏结构稳定性或结合能力，进而影响互作。

3.2. 二级结构对于蛋白质结构的影响

蛋白质的二级结构作为连接一级序列与高级构象的关键环节，主要描述其局部折叠模式[9]。典型的二级结构单元(如 α -螺旋和 β -折叠)通过局部空间排列形成稳定的支撑框架和相互作用界面。这些结构的稳定性主要依赖于肽链中氨基酸残基之间的非共价相互作用，包括氢键、静电吸引力以及范德华力，这些弱相互作用共同维持了二级结构的精确构象[10]。二级结构的稳定性也影响结合动力学，柔性区域可能参与诱导契合。

3.3. 三级结构对于蛋白质结构的影响

三级结构通过三维折叠形成精确的表面拓扑结构，确保互作的特异性[11]。例如，抗体的互补决定区(CDR)与抗原表位的空间互补，或变构蛋白因配体结合引发的远端构象变化(如 G 蛋白偶联受体的激活)。结构域的柔性或刚性影响结合亲和力和可调节性。

3.4. 四级结构对于蛋白质结构的影响

四级结构涉及多亚基的组装，形成功能复合物(如血红蛋白、蛋白酶体)。亚基间的排布和动态调控(如协同效应、亚基交换)扩展了蛋白质互作的多样性和可塑性。例如，泛素连接酶复合物的重组可改变底物特异性，而信号转导中的多蛋白复合物(如 NF- κ B 通路)通过亚基组合实现精确调控。

4. 模型预测材料与方法

4.1. 数据集

4.1.1. 数据预处理流程

为了确保数据的质量和一致性，为了确保数据的质量和一致性，从 UniRef 数据库获取蛋白质序列数据，从 BioGRID 数据库获取 PPI 数据，并对这些数据进行了严格的预处理。具体步骤如下：首先，对蛋白质序列进行冗余信息的剔除，通过设置序列相似性阈值为 0.9，移除了高度相似的序列，以避免过拟合。其次，对 PPI 数据进行了清洗，去除了重复记录和不完整的记录，确保数据的可靠性。最后，通过标准化处理，将所有数据转换为统一的格式，例如将蛋白质序列长度统一调整为 1000 个氨基酸，以便后续的模式训练和验证。

4.1.2. 正负样本构建策略

在构建训练集时，我们采用了正负样本均衡的策略。正样本由已知的 PPI 数据构成，共收集了 5000 个正样本对。负样本则通过随机组合非互作的蛋白质对生成，同样生成了 5000 个负样本对。为了避免引入噪声，对负样本进行了严格的筛选，确保其与正样本在序列相似性和结构特征上具有可比性。具体来说，要求负样本对中蛋白质序列的相似性不超过 0.7，并且结构特征如二级结构组成相似度不超过 0.6。此外，还采用了数据增强技术，如序列变异(对序列进行 5% 的突变率处理)和结构扰动(对结构特征进行 10% 的随机扰动)，来扩充训练样本，以提高模型的泛化能力。

4.2. 深度学习模型构建

4.2.1. 网络架构创新点阐述

本研究的深度学习模型整合了卷积神经网络(CNN)与长短期记忆网络(LSTM)的结构优势。具体来说，我们设计了包含 3 个卷积层和 2 个 LSTM 层的网络结构。CNN 部分的卷积核大小分别为 3×3 、 5×5 和 7×7 ，滤波器数量分别为 64、128 和 256，步长为 1，激活函数采用 ReLU。LSTM 层的隐藏单元数分别为 256 和 512，采用 tanh 激活函数。在此基础上，我们引入了自注意力机制，以动态调整特征权重，突出关键的蛋白质互作位点。自注意力机制的具体实现方式为：计算查询向量(Query)和键向量(Key)的点积，得到注意力分数，然后通过 softmax 函数进行归一化，最后将归一化的注意力分数与值向量(Value)进行加权求和，得到加权后的特征表示。这种多尺度特征整合策略显著增强了模型对复杂蛋白质-蛋白质相互作用(PPI)模式的识别能力，使其能够更精确地预测蛋白质间的潜在互作关系。

4.2.2. 特征融合机制解析

为了充分利用蛋白质的多模态信息，我们设计了一种特征融合机制。首先，从蛋白质序列中提取氨基酸组成、二肽组成和进化信息等特征[12]，其中氨基酸组成特征的维度为 20，二肽组成特征的维度为 400，进化信息特征的维度为 20。其次，结合蛋白质结构信息(如残基接触图和表面电荷分布)，残基接触图特征的维度为 1000×1000 ，表面电荷分布特征的维度为 1000。然后，通过嵌入层将不同模态的数据映射到统一的特征空间，嵌入维度统一为 256。最后，通过注意力机制对融合后的特征进行加权，具体来说，对融合后的特征矩阵进行线性变换得到查询向量(Query)、键向量(Key)和值向量(Value)，计算注意力分数并进行归一化后，与值向量进行加权求和，从而增强关键特征的表达能力。

4.3. 模型性能评估

在模型评估过程中，我们选用准确率(Accuracy)、召回率(Recall)、F1 值、AUC 值、精确率(Precision)

和 Matthews 相关系数(MCC)作为关键指标来评估模型的预测性能[13]。这些指标从不同维度全面反映了模型的预测能力：准确率用于评估模型预测正确的比例，召回率反映了模型识别正样本的能力，精确率用于衡量模型预测正样本的准确性，F1 值通过平衡准确率和召回率，提供了两者的综合衡量指标，AUC 值用于评估模型区分正负样本的能力，MCC 则能更好地处理类别不平衡问题，对模型的整体性能进行综合评价。通过综合分析这些指标，我们可以全面评估模型在蛋白质 - 蛋白质相互作用(PPI)预测任务中的性能和有效性。

5. 模型预测结果

5.1. 模型性能评估

为了评估所提出的深度学习模型在蛋白质 - 蛋白质互作(PPI)预测中的性能，我们在多个独立的数据集上进行了严格的测试，并使用了一系列标准的评估指标，包括准确率(Accuracy)、召回率(Recall)、F1 值、AUC 值、精确率(Precision)和 Matthews 相关系数(MCC) [14]。实验结果显示，模型的预测准确率达到 85.6%，召回率为 88.2%，F1 值为 86.8%，AUC 值为 0.92，精确率为 87.4%，MCC 为 0.78。这表明模型能够正确预测大部分蛋白质对之间的互作状态，并且在识别正样本方面表现出色。通过引入精确率和 MCC 指标，我们更全面地评估了模型的性能，进一步验证了深度学习在 PPI 预测中的潜力。

5.2. 模型对比分析

为了进一步验证模型的有效性，我们进行了消融实验，验证不同模型组件的贡献。实验结果表明，当移除自注意力机制时，模型的准确率下降至 82.3%，召回率下降至 84.5%，F1 值下降至 83.4%，AUC 值下降至 0.88，精确率下降至 83.1%，MCC 下降至 0.72。这充分证明了自注意力机制在捕捉关键特征方面的重要性。此外，我们还将所提出的深度学习模型与几种现有的 PPI 预测方法进行了对比，包括传统的机器学习方法(支持向量机 SVM、随机森林 RF 等)和基于深度学习的其他模型(如 ProBID-Net、DeepRank2 等)，以及一些最新的 PPI 预测方法(如 DeepPPIISP、GraphPPIPred 等) [15]。从对比结果可以看出，本研究提出的深度学习模型在所有评估指标上均优于现有的方法，尤其是在 AUC 值和 MCC 上，表明模型在区分互作和非互作蛋白质对方面具有更高的准确性。

5.3. 模型泛化性能评估

为了全面评估模型的泛化能力，我们在多个独立数据集上实施了交叉评估。实验结果显示，模型在不同数据集上的表现具有一致性，准确率波动幅度不超过 2%，召回率波动幅度不超过 3%。例如，在数据集 A 上，模型的准确率为 85.2%，召回率为 87.8%；在数据集 B 上，准确率为 86.1%，召回率为 88.5%；在数据集 C 上，准确率为 84.9%，召回率为 87.3%。这一结果验证了模型在不同数据分布下的稳定性，表明其具备较强的泛化能力，能够可靠地应用于多种场景的蛋白质 - 蛋白质相互作用(PPI)预测任务中[16]。

6. 模型优劣性分析

6.1. 模型优势

本模型首次将蛋白质进化信息与三维结构特征进行多尺度融合，通过注意力机制动态捕捉序列-结构关联模式。具体来说，我们不仅整合了蛋白质序列的氨基酸组成、二肽组成和进化信息等特征，还结合了蛋白质结构的残基接触图和表面电荷分布等信息。通过引入自注意力机制，模型能够动态调整特征权重，突出关键的蛋白质互作位点。此外，模型的自注意力机制和卷积神经网络能够有效地捕捉蛋白质序列中的长期依赖关系和局部特征，从而提高了 PPI 预测的准确性[17]。

6.2. 模型的局限性

尽管模型在 PPI 预测中取得了显著的性能提升,但仍存在一些局限性。例如,模型对蛋白质的三维结构特征的利用存在一定的局限性,目前仅纳入了残基接触图和表面电荷分布等基本信息,而蛋白质的三维结构特征在 PPI 中扮演着至关重要的角色。这种限制可能影响模型对某些复杂互动模式的识别能力。此外,模型的训练依赖于大量的已知 PPI 数据,对于一些数据稀缺的蛋白质家族,模型的性能可能会受到影响。

6.3. 未来工作方向

为进一步提高模型的性能,未来研究可以考虑以下几个方向:

1) 结合更全面的蛋白质三维结构信息:将蛋白质的三维结构数据(如残基接触图、表面电荷分布等)纳入模型,以更全面地捕捉 PPI 的关键特征。

2) 多模态数据融合:整合蛋白质序列、结构、基因表达和代谢数据等多种生物信息学数据,进一步提升模型的预测能力[18]。

3) 预训练模型的应用:利用预训练模型(如 ESM-2、MP-BERT)进行自监督学习,以提高模型对蛋白质序列的理解能力。

4) 数据增强:通过数据增强技术(如序列变异、结构扰动)生成更多的训练样本,提高模型的泛化能力。

7. 蛋白质互作研究的未来展望与技术革新

7.1. 技术创新

7.1.1. 氨基酸序列编码方法的创新

在蛋白质序列分析领域,我们提出了一种基于氨基酸序列相关系数的新型编码方法。该方法通过量化氨基酸序列内部长程相互作用以及序列间的协同进化关系,为蛋白质功能预测和互作研究提供了新的视角。具体而言,该编码方式不仅捕捉了单个蛋白质内部的结构特征,还特别关注了蛋白质序列内部长程相互作用的动态特性,以及不同蛋白质序列之间的协同进化模式。这种创新的编码策略能够更全面地揭示蛋白质的生物学功能及其在复杂分子网络中的角色,为蛋白质-蛋白质相互作用(PPI)的预测和分析提供了更强大的工具[19]。

这种创新方法旨在通过量化氨基酸间的相关性来捕捉蛋白质功能及其在细胞中的角色,为研究蛋白质间的互作提供了新的视角。通过对大量蛋白质序列数据进行分析,这种方法能够揭示那些对蛋白质结构和功能至关重要的关键氨基酸残基,进而为设计更有效的药物靶点提供理论支持。

7.1.2. 免疫共沉淀探新互作

现今免疫共沉淀(Co-Immunoprecipitation, Co-IP)已成为测定特定蛋白(通常称为诱饵蛋白)是否结合新蛋白质的方法之一[20]。此技术的基本流程是将含有目标蛋白的溶液过柱,柱中预装有特异性抗体。当目的蛋白流经此柱时,它会与预先固定的抗体结合,并捕获任何直接或间接与其互作的猎物蛋白。这些猎物蛋白可能结合于琼脂糖珠上或 GST 本身。之后,通过洗脱步骤将这些复合物分离出来,并利用 SDS-PAGE(十二烷基硫酸钠聚丙烯酰胺凝胶电泳)进行分析。这一步骤允许研究人员观察到与目标蛋白互作的所有蛋白质,从而帮助识别潜在的新伙伴或验证已知的互作。

7.1.3. 蛋白芯片技术的高通量应用

作为一种高通量、快速、灵敏、微型化及自动化的蛋白质分析手段,蛋白质芯片近年来在生物医学

研究中展现出巨大潜力。其核心方法涉及依据预定布局将大量蛋白质分子固定于固态载体，生成高密度微阵列。这些固定化探针蛋白可特异性结合样品中的目标蛋白，通过专用设备实现目标蛋白的定性与定量分析[21]。该技术借鉴了 DNA 芯片的设计思路，但针对蛋白质分子的特性进行了优化，提升了固定稳定性和检测灵敏度。通过构建高密度探针蛋白点阵，研究人员能够同时评估大量蛋白质样本的表达水平，显著提升研究效率与数据准确性。

蛋白质芯片的广泛应用涵盖基础研究、临床诊断及个性化医疗领域。在基础研究中，它助力于蛋白质功能鉴定和互作网络解析；在临床诊断中，可用于疾病标志物筛查和病情监测；在个性化医疗中，则为精准治疗方案的设计提供关键数据支持。这种技术进步不仅加速了疾病机制的理解，还为创新疗法的开发提供了强有力的工具，推动了转化医学的发展进程。

7.2. 多组学整合研究

7.2.1. 新技术助力蛋白多肽研究

随着蛋白质组学、多肽组学以及高分辨质谱技术的持续进步，蛋白多肽类成分的高通量分析迎来了新的突破。这些前沿技术不仅为蛋白多肽的鉴定和定量提供了高效手段，还为深入解析蛋白质功能及疾病机制提供了强大的工具[22]。通过整合多组学数据，研究人员能够更全面地理解生物系统的复杂性，从而加速新疗法的开发和个性化医疗的实施。这些前沿技术不仅促进了对动物药中蛋白多肽类成分的研究，还开辟了理解和分析复杂生物系统的新途径。通过这些先进的工具，研究人员能够更深入地探索生物体内复杂的分子网络，从而为药物开发和疾病治疗提供宝贵的信息。

7.2.2. 系统生物学整合多层次信息揭示生物系统全景

系统生物学通过融合基因组学、蛋白质组学和代谢组学等多层次的生物信息，全面解析生物体系的整体状态。该方法将单一层面的互作、代谢途径和调控网络进行整合，构建出生物系统的全景图。这种方法非常适合在研究初期构建出一个轮廓图，为后续更为深入的探究奠定基础，并指明方向[23]。通过系统生物学的方法，研究人员能够更高效地识别潜在的治疗靶点，并为个性化医疗和精准医学的发展奠定基础。

7.2.3. 代谢组学蛋白质组学解析生命动态

代谢组学是研究生物体内代谢物变化规律的一门新兴技术，已被用来研究多种疾病机制。其中，基于细胞的代谢组学分析主要用来研究药物及基因引起的代谢重编程。蛋白质组学是研究生物体内大分子蛋白质变化规律的技术[24]。基于细胞的代谢组学分析专注于药物干预或基因修饰导致的代谢重编程，有助于精确描绘疾病状态下代谢路径的变化，发现潜在生物标志物和治疗靶点。蛋白质组学研究生物体内大分子蛋白质的变化规律，对于理解生命过程中的动态变化至关重要。这些技术共同深化了我们对生命科学的认识。

7.3. 临床应用拓展

7.3.1. GST Pull-Down 鉴定蛋白质互作

GST-pull down 技术是一种体外验证蛋白质间直接互作的实验方法，在生物医学研究中具有重要作用。其应用主要体现在两个方面：一是用于证实两种已知蛋白之间可能存在的互作，这对于理解细胞内复杂的信号传导路径和分子机制至关重要；二是寻找能够与已知蛋白质发生互作的未知蛋白，这有助于发现新的蛋白质功能和潜在的药物靶点[25]。通过 GST-pull down 实验，研究人员可以有效地鉴定出特定蛋白质之间的直接物理互作，为后续的功能研究提供坚实的基础。

7.3.2. 融合蛋白技术增强表达效率

融合蛋白技术在生物医学领域展现出广泛的应用前景，不仅在融合蛋白药物的开发中发挥着关键作用，还在蛋白质表达优化方面具有显著优势。具体而言，该技术通过将外源基因与融合蛋白的 C 端相连，避免了额外设计启动子(SD 序列)的必要性，同时 N 端的存在显著提升了外源基因的表达效率。此外，当外源基因与宿主蛋白的部分序列构建成融合基因并以融合蛋白形式表达时，能够有效降低宿主细胞蛋白酶对外源蛋白的降解风险，从而显著提高目标蛋白的产量和稳定性[26]。这些特点使得融合蛋白技术成为现代生物学研究和生物技术应用中不可或缺的工具，无论是在基础科研还是在医药产品开发上都显示出巨大的潜力和价值。通过利用这一技术，科学家们能够更有效地生产重组蛋白，加速新药研发进程，并深入探索生命科学中的复杂问题。

8. 结论

本研究提出了一种基于深度学习的 PPI 预测模型，通过整合蛋白质序列位点和结构信息，实现了高效的 PPI 预测[27]。实验结果表明，该模型在多个数据集上表现出色，优于现有的传统方法和其他深度学习模型，展现出强大的预测能力和优异的泛化性能。该模型的成功应用不仅为 PPI 预测提供了新的工具，也为深入理解蛋白质之间的互作机制、细胞内复杂的信号网络以及疾病发生发展的分子基础提供了重要的理论支持。在实际应用中，这一模型能够帮助生物学家更高效地识别潜在的蛋白质互作关系，从而加速疾病机制的解析和药物靶点的发现。同时，我们也意识到模型仍存在一些局限性。如，模型的训练依赖于较多的已知 PPI 数据，对于一些数据稀缺的蛋白质家族，模型的性能可能会受到影响[28]。

在蛋白质互作预测这一领域，随着技术的不断进步，从早期基于序列比对和简单机器学习的方法，到如今深度学习模型的广泛应用，研究手段日益丰富且高效[29]。本研究提出的模型正是这一发展进程中的重要成果，它不仅在预测性能上取得了显著提升，更为未来的研究提供了新的思路 and 方向。通过这些改进和拓展，我们期待该模型能够在生物医学研究和药物开发中发挥更大的作用，为揭示生命奥秘和改善人类健康贡献更多的力量。

参考文献

- [1] 侯天云, 陆小鹏, 朱卫国. 蛋白质翻译后修饰在蛋白质-蛋白质相互作用中的调控作用[J]. 科学通报, 2017, 62(8): 759-769.
- [2] 潘巍峻. Wnt 信号转导的分子机制及其生物学功能研究[D]: [博士学位论文]. 上海: 中国科学院研究生院(上海生命科学研究院), 2004.
- [3] 李寅珠. 基于蛋白质相互作用网络的代谢 pathway 预测[D]: [硕士学位论文]. 天津: 天津师范大学, 2012.
- [4] 杜静静, 冯书风, 马晨. 基因表达调控的研究发展[J]. 今日畜牧兽医, 2018, 34(3): 64-65.
- [5] 沈珂, 吴群英, 蒋晓山. 双分子荧光互补技术在蛋白质相互作用研究中的应用[J]. 中国医学创新, 2015, 12(12): 132-135.
- [6] 陈谋通, 刘建军. 蛋白质相互作用的研究方法[J]. 生物技术通报, 2009(1): 50-54, 68.
- [7] 余晶梅. 荧光探针法和疏水相互作用层析法分析蛋白表面疏水性[D]: [硕士学位论文]. 杭州: 浙江大学, 2014.
- [8] 柴家乐. 基于静电相互作用的蛋清蛋白-多糖多尺度复合体系构建及应用研究[D]: [硕士学位论文]. 长春: 吉林大学, 2022.
- [9] 张安胜. 深度学习在蛋白质二级结构预测中的应用研究[D]: [硕士学位论文]. 合肥: 安徽大学, 2015.
- [10] 刘野. 基于分子动力学模拟研究蛋白质二级结构变化对其活性的影响[D]: [博士学位论文]. 长春: 吉林大学, 2019.
- [11] “阿尔法折叠”精准预测蛋白质三维结构[J]. 中国科学探险, 2021(3): 62.
- [12] 曹汉童. 基于深度学习的蛋白质相互作用分类预测研究[D]: [硕士学位论文]. 无锡: 江南大学, 2023.

-
- [13] 郭菲. 蛋白质相互作用预测方法及应用[Z]. 天津: 天津大学, 2020-08-24.
- [14] 靳晓宏. 基于深度学习预测蛋白质相互作用的研究[D]: [硕士学位论文]. 南宁: 广西民族大学, 2024.
- [15] 刘凡殊. 基于机器学习的蛋白质相互作用预测研究[D]: [硕士学位论文]. 烟台: 山东工商学院, 2022.
- [16] 张龙. 基于深度学习的蛋白质相互作用预测研究[D]: [硕士学位论文]. 重庆: 西南大学, 2019.
- [17] 叶明德. 集成神经网络与新编码方法预测蛋白质相互作用[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2021.
- [18] 李丽萍. 基于生物信息学的饲用植物与病原物蛋白质相互作用网络构建及应用[D]: [博士学位论文]. 乌鲁木齐: 新疆农业大学, 2022.
- [19] 耿昊. 基于 Transformer 的蛋白质间相互作用预测方法研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2024.
- [20] 郭睿, 刘全忠. 蛋白质相互作用研究技术的新进展[J]. 天津医科大学学报, 2015, 21(6): 542-544.
- [21] 王兵. 蛋白质相互作用及其位点的预测方法研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2006.
- [22] 杨彬, 高文远, 张艳军. 基于转录组学-蛋白质组学-多肽组学整合关联分析策略的动物药蛋白多肽类成分研究思路及方法[J]. 中草药, 2019, 50(5): 1033-1038.
- [23] 赵红霞. 蛋白质组学-代谢组学整合策略用于血管性抑郁小鼠模型病理机制探索的研究[D]: [硕士学位论文]. 上海: 第二军医大学, 2017.
- [24] 颜敏, 刘静, 夏天, 等. 细胞蛋白质组学和代谢组学整合策略表征散斑型 BTB/POZ 蛋白质突变调控的关键代谢通路[J]. 色谱, 2019, 37(8): 887-896.
- [25] 柴政斌, 张更林, 韩金祥. GST-pulldown 技术在蛋白质相互作用中的应用[J]. 中国生物制品学杂志, 2014, 27(10): 1354-1358.
- [26] 汪建华. 融合蛋白技术的应用[J]. 陕西医学杂志, 2004, (12): 1128-1129.
- [27] 许文俊. 面向蛋白质相互作用的表示学习方法研究[D]: [博士学位论文]. 合肥: 安徽农业大学, 2024.
- [28] 马文政. 基于特征选择与特征融合的蛋白质相互作用分类研究[D]: [硕士学位论文]. 济南: 济南大学, 2022.
- [29] 孙晶京. 基于序列的蛋白质相互作用预测方法研究进展[J]. 云南化工, 2020, 47(9): 13-14, 17.