

同行专家业内评价意见书编号: 20250860034

浙江工程师学院（浙江大学工程师学院） 同行专家业内评价意见书

姓名: 卞晴宇

学号: 22260233

申报工程师职称专业类别（领域）: 生物与医药

浙江工程师学院（浙江大学工程师学院）制

2025年03月13日

填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位+申报工程师职称专业类别(领域)4位+流水号3位，共11位。

一、个人申报

(一) 基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

1. 对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

卞晴宇同学在研究生期间系统学习了新药创制工程、高阶工程认知、产业技术前沿、蛋白质工程、优化算法等核心专业课程，广泛了解了生物医药方面的理论知识，并能够将这些理论知识与深度学习方法相结合，应用于蛋白互作预测与生物医学研究。她熟练使用蛋白数据库（UniProt、PDB、STRING、BioGRID、Pfam），能够进行蛋白序列特征提取、蛋白家族分类、蛋白功能预测，并利用生物大数据挖掘技术进行蛋白互作网络构建及功能富集分析。综合来看，卞晴宇同学在蛋白质结构、蛋白互作预测、计算蛋白质学、生物医药交叉研究等领域具备扎实的理论基础和技术能力，并能够结合前沿计算方法探索蛋白互作的分子机制，为生物信息学和精准医学研究提供技术支持。

2. 工程实践的经历(不少于200字)

在研究生期间积极参与蛋白互作预测模型的工程实现与应用，在深度学习模型开发、计算平台构建、数据处理优化等方面积累了丰富的工程实践经验，并推动了研究成果在生物信息学领域的落地应用。

在深度学习模型开发方面，我基于Transformer、注意力机制（Self-Attention, Co-Attention）、图神经网络（GNN）等技术，提出并优化了CAPPI（基于协同注意力机制的蛋白互作预测模型）。他不仅构建了高效的神经网络架构，还优化了数据预处理、特征提取、模型训练策略，使模型在大规模蛋白数据集上的计算效率和预测精度得到提升。

在计算平台的工程实现方面，我主导开发了DL Lab 在线预测平台，实现了蛋白-蛋白互作用（PPI）预测、蛋白互作位点预测（PPI-site）等功能。平台采用Python + Flask + PyTorch 进行后端开发，并结合前端可视化技术（如D3.js、ECharts），提供了蛋白结构可视化、交互式 PPI预测结果展示等功能，使得模型能够直接服务于生物医学研究人员。此外，为提升计算效，我对模型进行了GPU 并行计算优化，大幅提升了大规模数据处理的速度。

在大规模生物数据处理方面，我熟练使用生物数据库（UniProt、PDB、STRING、BioGRID），并基于Pandas、NumPy、Scikit-learn 进行蛋白序列和蛋白互作数据的高效处理。他优化了蛋白特征编码流程，使 CAPPI 能够高效处理长序列蛋白数据，并利用分布式计算（如 PySpark）加速超大规模数据集的训练，确保了模型在高通量生物信息学任务中的可用性。

此外，我在模型优化和可解释性分析方面，结合了注意力权重可视化、蛋白互作网络构建等方法，使得模型不仅能够准确预测 PPI，还能直观展现蛋白相互作用的关键残基，提高了模型的生物学可解释性。

综合来看，卞晴宇同学在深度学习模型开发、计算平台构建、大规模数据处理、GPU 并行计算优化等方面均积累了丰富的工程实践经验，并推动了 CAPPI 及 DL Lab 平台的落地应用，为蛋白互作研究提供了强有力的计算支持。

3. 在实际工作中综合运用所学知识解决复杂工程问题的案例(不少于1000字)

蛋白-蛋白相互作用 (PPI) 是生命活动的基础, 对于理解细胞信号传导、疾病发生及药物研发具有重要意义。然而, 由于实验验证 PPI 方法如酵母双杂交Y2H、亲和纯化等成本高、周期长, 如何利用计算方法进行高效、准确的 PPI预测成为生物信息学领域的重要研究方向。

在研究生阶段, 我基于深度学习方法, 构建了一种基于协同注意力机制 (Co-Attention) 的 PPI 预测模型 CAPPI (Co-Attention Protein-Protein Interaction), 并搭建了在线计算平台 DL Lab, 为生物医学研究人员提供高效、便捷的蛋白互作预测服务。该工程项目涉及深度学习模型构建、数据预处理优化、大规模计算平台部署、GPU并行计算加速、蛋白结构可视化等多个方面, 具有较强的工程复杂性。

在构建 PPI 预测模型过程中, 我们面临了多个复杂工程问题, 包括:

(1) 大规模蛋白质数据的高效处理

蛋白质序列数据通常由 UniProt、PDB、STRING、BioGRID等数据库提供, 数据量巨大, 且不同数据库的数据格式不统一。例如UniProt提供蛋白序列数据, 但不同物种的注释信息不同。PDB提供蛋白质三维结构数据, 但数据不全, 且需要标准化解析。STRING 和 BioGRID提供蛋白互作数据, 但不同数据来源的置信度不同。如何高效整合这些异构数据源, 并进行统一编码, 是 PPI 预测的基础挑战。

(2) 蛋白序列的深度学习建模

蛋白质的一级序列包含数百至上千个氨基酸, 传统的 CNN 和 RNN由于受限于局部感受野和长程依赖问题, 难以捕捉蛋白质远程残基的相互作用。此外, 实际的蛋白相互作用过程通常涉及多个蛋白质协同参与, 但现有模型多依赖于单一蛋白链进行特征提取, 这与实际的生物过程不符, 限制模型的可解释性和应用。同时, 复杂模型结构带来的高计算成本和部署难度, 阻碍了其他领域研究者的广泛应用。

为解决这些工程问题, 我运用所学深度学习知识与生物医药知识设计并搭建全新的基于序列的PPI预测模型。

首先, 模型嵌入基于ELMo和大规模蛋白序列数据集的预训练语言模型动态地进行特征编码, 利用ELMo可捕捉上下文长距离依赖的优势, 加强对远距离残基间关系的表征。

其次, 基于独特的分支结构, 支持完成PPI及PPI-site两类任务。

同时, 模型以成对的蛋白序列作为输入, 基于全新的协同注意力结构有效整合并学习参与互作的多个蛋白质间的复杂协同关系, 充分考虑到PPI过程中蛋白间的相互影响, 贴合实际生物过程。设计HieDPA模块, 通过灵活和层次化的注意力机制来聚合蛋白质序列的特征, 克服传统特征聚合方式难以捕捉长程依赖的局限性。

此外, 从数据多样性、泛用性和高质量等多个维度构建了多个数据集, 并通过调整损失函数权重来补偿数据不平衡问题。同时, 采用修改测试阈值的方法平衡模型的查准率和查全率, 进一步优化预测性能。

最后, 在多个数据集上与现有基准模型进行性能对比测试, 并在跨物种及低一致性数据集上进行鲁棒性实验。此外, 通过消融实验和可解释性分析, 验证模型的优异性能, 证明其能够适应不同任务场景的需求, 推动PPI预测领域的进一步发展。

(二) 取得的业绩(代表作)【限填3项, 须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等)供核实, 并提供复印件一份】

1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】

成果名称	成果类别 [含论文、授权专利(含发明专利申请)、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/ 授权或申 请时间等	刊物名称 /专利授权 或申请号等	本人 排名/ 总人 数	备注
PPI-CoAttNet: A Web Server for Protein-Protein Interaction Tasks Using a Coattention Model	核心期刊	2025年02月06日	Journal of Chemical Information and Modeling	1/12	SCI期刊收录

2. 其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】

(三) 在校期间课程、专业实践训练及学位论文相关情况

课程成绩情况	按课程学分核算的平均成绩: 84 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间: 1.5 年(要求1年及以上) 考核成绩: 81 分
本人承诺	
个人声明: 本人上述所填资料均为真实有效, 如有虚假, 愿承担一切责任, 特此声明!	
申报人签名: 卞晴宇	

二、日常表现考核评价及申报材料审核公示结果

日常表现考核评价	非定向生由德育导师考核评价、定向生由所在工作单位考核评价: ☒优秀 ☐良好 ☐合格 ☐不合格 德育导师/定向生所在工作单位分管领导签字（公章）: 14 展 2025 年 3 月 30 日
申报材料审核公示	根据评审条件，工程师学院已对申报人员进行材料审核（学位课程成绩、专业实践训练时间及考核、学位论文、代表作等情况），并将符合要求的申报材料在学院网站公示不少于5个工作日，具体公示结果如下： ☐通过 ☐不通过（具体原因： ） 工程师学院教学管理办公室审核签字（公章）: 年 月 日

浙 江 大 学 研 究 生 院
攻读硕士学位研究生成绩表

学号：22260233		姓名：卞晴宇		性别：女		学院：工程师学院		专业：生物与医药			学制：2.5年				
毕业时最低应获：24.0学分				已获得：26.0学分				入学年月：2022-09			毕业年月：				
学位证书号：				毕业证书号：				授予学位：							
学习时间		课程名称		备注	学分	成绩	课程性质	学习时间		课程名称		备注	学分	成绩	课程性质
2021-2022学年春季学期		研究生英语			2.0	免修	专业学位课	2022-2023学年冬季学期		产业技术发展前沿			1.5	90	专业学位课
2021-2022学年春季学期		研究生英语基础技能			1.0	免修	公共学位课	2022-2023学年春季学期		自然辩证法概论			1.0	78	专业学位课
2022-2023学年秋季学期		工程技术创新前沿			1.5	88	专业学位课	2022-2023学年春季学期		蛋白质工程			2.0	91	专业选修课
2022-2023学年秋季学期		新时代中国特色社会主义思想理论与实践			2.0	89	专业学位课	2022-2023学年春夏学期		高阶工程认知实践			3.0	86	专业学位课
2022-2023学年秋冬学期		工程伦理			2.0	84	专业学位课	2022-2023学年夏季学期		药品创制工程实例			2.0	86	专业学位课
2022-2023学年冬季学期		新药发现理论与实践			2.0	87	专业学位课	2022-2023学年春夏学期		优化算法			3.0	71	专业选修课
2022-2023学年秋冬学期		研究生论文写作指导			1.0	79	专业选修课			硕士生读书报告			2.0	通过	

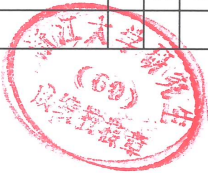
说明：1. 研究生课程按三种方法计分：百分制，两级制（通过、不通过），五级制（优、良、中、及格、不及格）。

2. 备注中“*”表示重修课程。

学院成绩校核章：

成绩校核人：张梦依

打印日期：2025-03-20



PPI-CoAttNet: A Web Server for Protein–Protein Interaction Tasks Using a Coattention Model

Qingyu Bian,[#] Zheyuan Shen,[#] Jian Gao,[#] Liteng Shen, Yang Lu, Qingnan Zhang, Roufen Chen, Donghang Xu, Tao Liu, Jinxin Che,^{*} Yan Lu,^{*} and Xiaowu Dong^{*}



Cite This: *J. Chem. Inf. Model.* 2025, 65, 461–471



Read Online

ACCESS |



Metrics & More

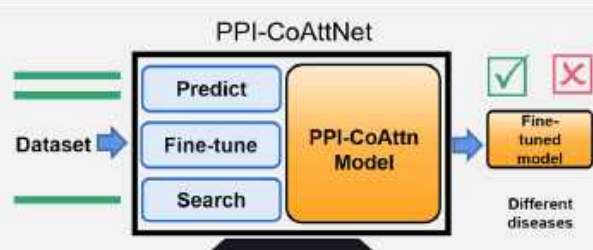


Article Recommendations



Supporting Information

ABSTRACT: Predicting protein–protein interactions (PPIs) is crucial for advancing drug discovery. Despite the proposal of numerous advanced computational methods, these approaches often suffer from poor usability for biologists and lack generalization. In this study, we designed a deep learning model based on a coattention mechanism that was capable of both PPI and site prediction and used this model as the foundation for PPI-CoAttNet, a user-friendly, multifunctional web server for PPI prediction. This platform provides comprehensive services for online PPI model training, PPI and site prediction, and prediction of interactions with proteins associated with highly prevalent cancers. In our *Homo sapiens* test set for PPI prediction, PPI-CoAttNet achieved an AUC of 0.9841 and an F1 score of 0.9440, outperforming most state-of-the-art models. Additionally, these results are generated in real time, delivering outcomes within minutes. We also evaluated PPI-CoAttNet for downstream tasks, including novel E3 ligase scoring, demonstrating outstanding accuracy. We believe that this tool will empower researchers, especially those without computational expertise, to leverage AI for accelerating drug development.



INTRODUCTION

Protein–protein interactions (PPIs) form the foundational basis of cellular structural organization and functional coordination.¹ These interactions are crucial in numerous biological processes, including signal transduction, cellular communication, and metabolic pathways.² In-depth research into PPIs not only provides profound insights into cellular mechanisms but also opens new avenues for the identification of novel therapeutic targets and the development of drugs designed to modulate these interactions.³

Traditional experimental methods are commonly used to assess protein–protein binding and identify the critical amino acids at the binding interface. Although these techniques have proven to be relatively accurate, they still have certain limitations.⁴ X-ray crystallography provides high-resolution structures of protein complexes but requires high-quality crystals, which can be difficult to obtain for some proteins. Cryogenic electron microscopy allows for the study of large and flexible protein complexes without the need for crystallization, but it can be costly and typically provides lower resolution for smaller complexes. Alanine scanning⁵ identifies key residues in PPIs by substituting amino acids with alanine, though it may not capture all types of interactions, particularly those involving subtle conformational changes. Consequently, there is a need for computational tools to facilitate the automated prediction and understanding of interactions between biomolecules.⁶

Various databases, such as STRING,⁷ BioGRID,⁸ and the recently developed PPI3D,⁹ have been constructed to facilitate PPI studies by inferring protein–protein interactions based on sequence homology and structural similarity. However, these inferences rely on existing data and do not fully capture the underlying relationships between interacting proteins, leading to limitations in predicting interactions for unknown proteins.

In recent years, tools such as AlphaFold3¹⁰ and RoseTTA-Fold-All-Atom¹¹ have been developed for the structural prediction of multiple protein complexes. However, obtaining the structures of complexes using these tools requires significant computational resources and extended computation times. Moreover, these tools often only provide the predicted interaction structures without information about whether interactions occur or details about interacting amino acids. Additionally, some computational models for PPI prediction, including deep-learning-based methods using protein sequences, like PIPR,¹² GNN-PPI,¹³ HN-PPISP,¹⁴ and DeepPPISP,¹⁵ and methods based on protein 3D structures, such as Struct2Graph¹⁶ and RGN,¹⁷ often focus on single-chain

Received: July 31, 2024

Revised: December 14, 2024

Accepted: December 16, 2024

Published: January 6, 2025



经检索《Web of Science》和《Journal Citation Reports (JCR)》数据库,《Science Citation Index Expanded (SCI-EXPANDED)》收录论文及其期刊影响因子、分区情况如下。(检索时间:2025年3月13日)

第1条,共1条

标题:PPI-CoAttNet: A Web Server for Protein-Protein Interaction Tasks Using a Coattention Model

作者:Bian, QY(Bian, Qingyu);Shen, ZY(Shen, Zheyuan);Gao, J(Gao, Jian);Shen, LT(Shen, Liteng);Lu, Y(Lu, Yang);Zhang, QN(Zhang, Qingnan);Chen, RF(Chen, Roufen);Xu, DH(Xu, Donghang);Liu, T(Liu, Tao);Che, JX(Che, Jinxin);Lu, Y(Lu, Yan);Dong, XW(Dong, Xiaowu);

来源出版物:JOURNAL OF CHEMICAL INFORMATION AND MODELING 卷:65 期:2 页:461-471 提前访问日期:JAN 2025 DOI:10.1021/acs.jcim.4c01365 出版年:JAN 6 2025

入藏号:WOS:001391440300001

文献类型:Article

地址:

[Bian, Qingyu; Shen, Zheyuan; Gao, Jian; Shen, Liteng; Lu, Yang; Zhang, Qingnan; Chen, Roufen; Liu, Tao; Che, Jinxin; Dong, Xiaowu] Zhejiang Univ, Coll Pharmaceut Sci, Hangzhou 310058, Peoples R China.

[Xu, Donghang; Lu, Yan; Dong, Xiaowu] Zhejiang Univ, Affiliated Hosp 2, Dept Pharm, Sch Med, Hangzhou 310058, Peoples R China.

[Bian, Qingyu; Shen, Zheyuan; Gao, Jian; Shen, Liteng; Lu, Yang; Chen, Roufen; Liu, Tao; Che, Jinxin; Dong, Xiaowu] Zhejiang Univ, Innovat Inst Artificial Intelligence Med, Hangzhou 310058, Peoples R China.

通讯作者地址:

Che, JX; Dong, XW (corresponding author), Zhejiang Univ, Coll Pharmaceut Sci, Hangzhou 310058, Peoples R China.; Lu, Y; Dong, XW (corresponding author), Zhejiang Univ, Affiliated Hosp 2, Dept Pharm, Sch Med, Hangzhou 310058, Peoples R China.; Che, JX; Dong, XW (corresponding author), Zhejiang Univ, Innovat Inst Artificial Intelligence Med, Hangzhou 310058, Peoples R China.

电子邮件地址:chejx@zju.edu.cn; 2504159@zju.edu.cn; dongxw@zju.edu.cn

IDS号:T7M3S

ISSN:1549-9596

eISSN:1549-960X

期刊《Journal of Chemical Information and Modeling》2023年的影响因子为5.7,五年影响因子为5.9。

期刊《Journal of Chemical Information and Modeling》2023年的JCR分区情况为:

Edition	JCR® 类别	类别中的排序	JCR 分区
SCIE	CHEMISTRY, MEDICINAL	10/72	Q1
SCIE	CHEMISTRY, MULTIDISCIPLINARY	61/231	Q2
SCIE	COMPUTER SCIENCE, INFORMATION SYSTEMS	34/250	Q1
SCIE	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	28/170	Q1

注:

1. 期刊影响因子及分区情况最新数据以 JCR 数据库最新数据为准。
2. 以上检索结果来自 CALIS 查收查引系统。
3. 以上检索结果均得到委托人及被检索作者的确认。

