

同行专家业内评价意见书编号： 20250860058

附件1

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： 张忆雪

学号： 22260241

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05月27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)

在人工智能与药物研发的交叉领域，我建立了扎实的理论基础与系统的专业知识体系。

1.

基础及专业知识：我深入掌握药物化学、计算生物学和机器学习三大核心学科的理论框架。在数学与计算科学方面，具备概率统计、最优化方法、分子动力学模拟等基础理论能力，能运用深度学习(如GNN、Transformer)、化学信息学(如分子描述符计算、QSAR建模)等技术解决药学问题。熟悉药物研发全流程中的关键ADMET性质(吸收、分布、代谢、排泄、毒性)及其计算预测方法，掌握分子表征学习、特征工程和模型评估的核心技术。同时，了解医药经济学、项目管理等相关人文与管理知识，确保技术方案符合实际研发需求。

2.

行业知识：我持续关注国内外AI制药领域的技术发展，熟悉行业主流技术(如虚拟筛选、de novo分子设计)和新兴趋势(如多模态学习、生成式AI在药物发现中的应用)。掌握药物研发的标准流程(如临床前研究规范)、行业数据标准(如FAIR原则)、以及监管要求(如FDA/EMA对计算模型的验证标准)，能够结合行业最佳实践优化技术方案。

3.

默会性工程知识：通过长期的专业实践，我积累了丰富的工程经验，包括(1)针对药物数据的高噪声、小样本特点，灵活运用数据增强、迁移学习等策略提升模型鲁棒性；(2)在模型开发中平衡计算效率与预测精度，确保方法在真实场景中的可用性；(3)结合药学专家的先验知识优化算法设计，增强模型的可解释性和可信度等。

2. 工程实践的经历(不少于200字)

在碳硅智慧科技有限公司担任研发实习生期间，我深度参与了AI驱动的成药性预测模型研发工作，积累了扎实的工程实践经验，涵盖数据构建、算法优化、模型开发及实际应用等关键环节。

1.

系统性数据工程与成药性数据集构建：为构建全面、准确的成药性预测模型，我参与了大规模、多源异构数据的收集、清洗与标准化工作。基于PubChem、ChEMBL、DrugBank等公开数据库，并结合行业合作数据，系统整合了化合物的物理化学性质(如溶解度、logP、pKa)及ADMET(吸收、分布、代谢、排泄、毒性)参数。在数据预处理阶段，我采用数据去重、异常值检测、一致性校验等方法，确保数据的高质量和可靠性，最终构建了结构化的成药性数据集，为后续建模奠定坚实基础。

2.

机器学习模型开发与优化：在模型构建阶段，我探索了多种机器学习(如随机森林、XGBoost)和深度学习(如GNN、Transformer)算法，结合分子指纹(Morgan Fingerprint)、分子描述符(RDKit计算)及图神经网络(GNN)表征，优化了成药性预测任务的特征工程流程。通过交叉验证、超参数调优、模型融合等技术，提升了模型的泛化能力，使其在溶解度、渗透性、肝毒性等关键ADMET指标上表现优异。此外，我引入了SHAP值分析、特征重要性排序等方法，增强模型的可解释性，便于药学专家理解预测逻辑。

3.

工程化落地与行业应用：为推进模型的实际应用，我参与了模型部署与性能优化工作，利用Flask/Django框架搭建了轻量级预测服务，并优化计算效率以适应大规模分子库的快速筛选

。同时，结合药物化学知识，对预测结果进行后处理分析，确保其符合生物合理性。该模型已成功应用于虚拟药物筛选，显著提升了早期药物发现的效率，并降低了实验验证成本。

4.

团队协作与跨领域知识融合：在项目推进过程中，我与药物化学家、计算生物学家、软件开发工程师紧密协作，深入理解药物研发的实际需求，确保技术方案与行业标准（如FAIR数据原则、OECD

QSAR建模准则）相契合。通过这一实践，我不仅提升了AI+药物研发的全流程工程能力，也深化了对药物发现复杂性问题的理解，为未来更深入的科研与工程应用积累了宝贵经验。

此次实习经历使我全面掌握了AI制药领域的工程实践方法，并培养了数据驱动、模型优化、跨学科协作的综合能力，为后续的技术研发奠定了坚实基础。

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）

在项目实践过程中，我们团队始终聚焦于将人工智能前沿技术与药物研发实际需求深度融合，通过系统性创新取得了显著的技术突破和实用价值。在项目初期，我们面临三大技术挑战：实验数据分散且标准不一、分子表征方法单一、现有预测模型可解释性能力不足。针对这些难题，我和团队开展了系统性攻关。

在构建智能化成药性预测体系的过程中，我们首先系统整合了来自PubChem、ChEMBL等国际权威数据库的化合物数据，同时通过与行业伙伴的战略合作，构建了覆盖溶解度、渗透性、代谢稳定性等关键指标的ADMET数据集。为确保数据质量，我们创新性地开发了多级数据清洗流程，包括基于化学规则的异常值检测、结构相似性去重以及实验数据一致性校验，最终形成了超过10万条高质量数据点的标准化数据集，为模型开发奠定了坚实基础。

在技术创新层面，我主导开发了一系列基于人工智能的药物吸收性质预测模型，模型聚焦于肠道渗透性和与P-

gp的相互作用两大关键因素。针对肠道渗透性预测，系统地比较了七种分子表征方法（RDKit、Mordred、ECFP4、MACCS、EState、ErG、Graph）和六种主流机器学习/深度学习算法（RF、SVM、XGBoost、HistGBM、GCN、MPNN）的性能，并探索了不同分子表征组合以及模型集成策略对预测结果的影响。研究表明，通过巧妙地融合不同分子表征，并采用Stacking集成学习方法（RF+XGBoost+HistGBM），可以显著提升肠道渗透性预测模型的性能。最终构建的模型在Caco-

2、MDCK和PAMPA数据集上均取得了优异的预测效果， R^2 分别达到0.807、0.750和0.755。在P-gp抑制剂和底物预测方面，创新性地提出了一种基于多模态对比学习的MC-

PGP模型。该模型通过注意力机制有效融合分子SMILES序列、分子指纹和分子图三类分子表征，实现了P-gp抑制剂与底物的高精度预测。具体而言，MC-PGP模型采用Bi-

LSTM结合注意力机制解析SMILES序列特征，通过FC层提取分子指纹特征，并利用GIN结合自动图对比学习策略捕捉分子图结构信息。实验结果表明，MC-PGP模型在P-

gp抑制剂和底物随机划分测试集上的AUC-

ROC分别达到0.939和0.940，显著优于现有先进算法。

为推动技术成果的产业化应用，我参与开发了AI驱动的一站式创新药物发现平台DrugFlow，DrugFlow涵盖苗头和先导分子发现及优化等药物研发关键环节，集成世界领先的原创分子对接、QSAR建模、成药性预测、分子生成、虚拟筛选算法，并提供用户友好的操作界面，旨在帮助药化专家更高效便捷地发现和设计具有开发潜力的成药分子。其中成药性预测模块是我主要参与的部分，在成药性预测模块中，DrugFollow平台提供了Inno-ADMET、ChemFH和Inno-SA三个模块，可对分子的成药性进行全面评估。其中Inno-

ADMET模块共支持76种物化性质、5种药化性质、21种成药性参数、27种毒性性质的系统性评价。目前该模块提供两种自主知识产权算法，分别为基于多任务图神经网络的MGA方法以及

基于多模态预训练语言模型的MERT方法，其中前者可结合自研的基于子结构掩盖的图神经网络可解释性方法SME识别关键子结构，为用户提供更为直观的模型解释，而后者则有望实现更为准确的性质预测。

针对工程化过程中的关键挑战，我们形成了一系列创新解决方案：在数据层面，建立了标准化的数据处理流程；在算法层面，开发了面向小样本学习的多表征融合策略；在应用层面，实现了模型预测与药物化学知识的有机结合。这些经验已固化为标准化的技术文档和操作规程，为后续更复杂的药物研发智能化项目提供了重要参考。我们的实践表明，通过深度整合人工智能技术与药物研发专业知识，完全可以在保证科学严谨性的前提下，显著提升研发效率和成功率，为医药行业的数字化转型提供了有价值的实践范例。通过本次实践，我深刻认识到解决复杂工程问题需要多学科知识的深度融合。未来，我们将进一步拓展模型覆盖的ADMET性质范围，探索生成式AI在分子优化中的应用，持续推动人工智能技术在药物研发中的深度应用。

(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
A Multimodal Contrastive Learning Framework for Predicting P-Glycoprotein Substrates and Inhibitors	国际期刊	2025年04月10日	Journal of Pharmaceutical Analysis	1/4	
一种基于多模态对比学习的P-糖蛋白抑制剂和底物预测方法	发明专利申请	2024年09月24日	申请号：CN 2024113337 28.6	2/6	

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： 85 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 年（要求1年及以上） 考核成绩： 84 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：张忆雪	

单位考核评价：
(章)：周展 1 年 5 月 18

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）：周晨 2023年5月18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：22260241	姓名：张忆雪	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：29.0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
2021-2022学年冬季学期	新药发现理论与实践		2.0	97	专业学位课
2021-2022学年春季学期	生物大分子模拟		2.0	88	专业选修课
2021-2022学年春夏学期	创业能力建设		2.0	84	跨专业课
2022-2023学年秋季学期	药物代谢学		2.0	88	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	91	公共学位课
2022-2023学年秋季学期	工程技术创新前沿		1.5	89	专业学位课
2022-2023学年秋冬学期	研究生论文写作指导		1.0	91	专业学位课
2022-2023学年冬季学期	产业技术发展前沿		1.5	93	专业学位课
2022-2023学年春季学期	数学建模		2.0	63	专业选修课
2022-2023学年春季学期	自然辩证法概论		1.0	80	公共学位课
2022-2023学年夏季学期	研究生英语基础技能		1.0	免修	公共学位课
2022-2023学年夏季学期	药品创制工程实例		2.0	92	专业学位课
2022-2023学年春夏学期	高阶工程认知实践		3.0	90	专业学位课
2022-2023学年夏季学期	研究生英语		2.0	免修	公共学位课
2023-2024学年秋冬学期	工程伦理		2.0	67	公共学位课
	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：2025-06-03

Journal Pre-proof

A Multimodal Contrastive Learning Framework for Predicting P-Glycoprotein Substrates and Inhibitors

Yixue Zhang, Jialu Wu, Yu Kang, Tingjun Hou



PII: S2095-1779(25)00130-3

DOI: <https://doi.org/10.1016/j.jpha.2025.101313>

Reference: JPHA 101313

To appear in: *Journal of Pharmaceutical Analysis*

Received Date: 24 September 2024

Revised Date: 7 March 2025

Accepted Date: 10 April 2025

Please cite this article as: Y. Zhang, J. Wu, Y. Kang, T. Hou, A Multimodal Contrastive Learning Framework for Predicting P-Glycoprotein Substrates and Inhibitors, *Journal of Pharmaceutical Analysis*, <https://doi.org/10.1016/j.jpha.2025.101313>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 The Author(s). Published by Elsevier B.V. on behalf of Xi'an Jiaotong University.



(21) 申请号 202411333728.6

G16C 20/90 (2019.01)

(22) 申请日 2024.09.24

G06N 3/084 (2023.01)

G06N 3/0985 (2023.01)

(71) 申请人 浙江大学

地址 310058 浙江省杭州市西湖区余杭塘路866号

(72) 发明人 康玉 张忆雪 侯廷军 吴佳璐
吴振兴 方美静(74) 专利代理机构 杭州求是专利事务有限公司
33200

专利代理师 陈升华

(51) Int. Cl.

G16C 20/30 (2019.01)

G16C 20/70 (2019.01)

G06N 3/045 (2023.01)

G06N 3/0442 (2023.01)

权利要求书2页 说明书15页 附图4页

(54) 发明名称

一种基于多模态对比学习的P-糖蛋白抑制剂和底物预测方法

(57) 摘要

本发明公开了一种基于多模态对比学习的P-糖蛋白抑制剂和底物预测方法,包括:构建P-糖蛋白抑制剂和底物的数据集;利用深度学习方法构建P-糖蛋白抑制剂和底物预测模型;输入到构建的P-糖蛋白抑制剂和底物预测模型中,将对比损失和分类损失联合得到总损失函数,通过总损失函数训练并优化P-糖蛋白抑制剂和底物预测模型的参数;将待预测的化合物数据输入到优化后的P-糖蛋白抑制剂和底物预测模型,输出最终的分类结果,判断化合物是否为P-糖蛋白抑制剂和底物。本发明提升了模型预测精度和泛化性能,并实现了模型的可解释性,为药物开发中的分子设计提供可靠的科学依据。

