

同行专家业内评价意见书编号: 20250860032

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: _____ 曹成

学号: _____ 22260234

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12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)

本人已掌握一定的药学基本理论，包括特定的药物作用机制、药物设计和开发。本人具备扎实的计算机科学知识，包括数据结构、算法、编程语言(如Python、Java等)以及数据库管理等。这些基础知识可以帮助我理解如何处理和分析大量的医学和药学数据，支持药物发现研究。已掌握人工智能的核心技术，包括机器学习、深度学习和数据挖掘等，尤其是深度学习技术，特别适合处理大规模、复杂的数据集，如图像、文本、分子结构数据等。我已充分理解这些技术并将其应用于药物发现、药物毒性预测等领域。

2. 工程实践的经历(不少于200字)

本人在中科院上海药物研究所负责基于分子多构象点云的发育毒性预测。旨在解决传统毒性预测方法难以捕捉分子多维度信息问题和无法融合多构象信息问题，进一步提高毒性预测的准确性。通过生成分子多范德华表面的点云数据，构建了基于分子多构象的深度学习模型，相比较传统基于二维的机器学习模型，该模型在预测精度上有较大的提高。同时为了克服黑盒效应，我们还开发了可视化模块，用于模型的可解释性，结果表明模型的可视化模块在探索化合物的作用机理方面可以起到辅助作用。

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)

研究内容：在分子毒性预测领域，传统的定量结构-

活性关系(QSAR)模型被广泛应用于建立分子结构与活性之间的相关性。然而，这些模型在本质上受到其依赖于分子的二维表示的限制，因此忽略了对毒性影响较大的分子相互作用的关键方面。分子毒性的产生往往源于分子与特定靶点(如蛋白质和核酸)的特定相互作用。在传统的基于二维分子特征方法中，对这些靶点的空间结构缺乏考虑，这可能会显著影响毒性预测的准确性。本研究旨在开发基于分子多构象的点云深度学习毒性预测方法，利用分子范德华表面的点云数据结构来表示分子，同时，基于数据驱动方式来融合分子多构象信息，以更准确地预测毒性。本课题的研究方法基于深度学习和点云表示。我们开发了一种多示例深度学习模型，用于从分子多构象的点云数据中预测毒性。这个模型将能够自动学习多构象之间的关系，从而更准确地进行毒性预测。该方法解决了传统毒性预测方法难以捕捉分子多构象的信息问题，通过引入点云表示，我们能够更自然地表示分子的多构象。同时也解决了传统方法无法融合多构象信息问题，基于数据驱动方式，自动学习融合多构象信息。这都有助于提高毒性预测的一致性和可靠性。

方案及技术路线：1) 数据收集和处理；2) 分子多构象点云数据生成；3) 模型构建、训练和评估；4) 结果分析和讨论。

该项目由我独自负责，目前取得的成果如下：1) 总共收集了510种斑马鱼发育毒性的化合物，用于使用MCPNET进行QSAR分析。2) 首先，我们应用了ETKDG方法，该方法将来自小分子晶体学数据的实验扭转角偏好整合到距离几何方法中，用于初始生成数据集中每种化合物的多种构象。其次，根据化合物的多构象生成范德华表面点云数据；3) 模型的构建。MCPNET架构主要由三个模块组成，一个基于分子范德华静电势点云数据提取信息向量的点云特征编码器，一个基于注意力的多构象池化机制，以及一个用于下游任务的多层感知器(MLP)。使用点云数据表示不同构象化合物的范德华静电势表面作为MCPNET的输入数据。采用点云特征编码器提取向量，丰富点云数据特征信息。然后使用基于注意力的多构象池化机制来聚合分子不同构象的特征。最后，通过MLP模块进行最终预测。4) MCPNET在分类任务上达到了85%的

高准确率，在回归任务上达到了0.29的低MSE值，证明了其有效性。此外，我们的模型优于其它传统模型，如RF，GCN等。并且对点云的解释性分析验证了MCPNET在探索化合物的作用机理方面可以起到辅助作用。

正在改进部分：1) 模型优化。调整模型架构和参数，进一步提升模型的性能。2) 数据集扩展。收集更多的发育毒性化合物和其它领域的数据集，将模型应用到不同领域。3) 开发综合性的毒性预测平台，集成模型功能和数据资源，为科研和工业界提供一站式服务。

(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
MCPNET: Development of an interpretable deep learning model based on multiple conformations of the compound for predicting developmental toxicity	TOP期刊	2024年03月01日	Computers in Biology and Medicine		
基于分子多构象的点云多示例深度学习毒性预测方法	发明专利申请	2024年01月09日	申请号: CN 2023112003 16		

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 年(要求1年及以上) 考核成绩: 83 分
本人承诺	
<p>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!</p> <p style="text-align: right;">申报人签名: 曹成</p>	

二、日常表现考核评价及申报材料审核公示结果

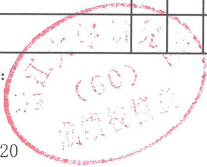
日常表现考核评价	<p>非定向生由德育导师考核评价、定向生由所在工作单位考核评价：</p> <p><input checked="" type="checkbox"/>优秀 <input type="checkbox"/>良好 <input type="checkbox"/>合格 <input type="checkbox"/>不合格</p> <p>德育导师/定向生所在工作单位分管领导签字（公章）： 2025年3月13日</p>
申报材料审核公示	<p>根据评审条件，工程师学院已对申报人员进行材料审核（学位课程成绩、专业实践训练时间及考核、学位论文、代表作等情况），并将符合要求的申报材料在学院网站公示不少于5个工作日，具体公示结果如下：</p> <p><input type="checkbox"/>通过 <input type="checkbox"/>不通过（具体原因： ）</p> <p>工程师学院教学管理办公室审核签字（公章）： 年 月 日</p>

浙 江 大 学 研 究 生 院
攻读硕士学位研究生成绩表

学号：22260234		姓名：曹成		性别：男		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：27.0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
2022-2023学年秋季学期		研究生英语基础技能			1.0	74	公共学位课	2022-2023学年春季学期		数学建模			2.0	68	专业选修课
2022-2023学年秋季学期		工程技术创新前沿			1.5	86	专业学位课	2022-2023学年春季学期		自然辩证法概论			1.0	75	公共学位课
2022-2023学年冬季学期		新时代中国特色社会主义思想理论与实践			2.0	88	公共学位课	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学年秋冬学期		工程伦理			2.0	95	公共学位课	2022-2023学年春夏学期		高阶工程认知实践			3.0	89	专业学位课
2022-2023学年冬季学期		产业技术发展前沿			1.5	84	专业学位课	2022-2023学年夏季学期		工程师创新创业思维			2.0	92	专业选修课
2022-2023学年冬季学期		新药发现理论与实践			2.0	89	专业学位课	2022-2023学年夏季学期		药物信息学			2.0	83	跨专业课
2022-2023学年秋冬学期		研究生论文写作指导			1.0	89	专业学位课			硕士生读书报告			2.0	通过	
2022-2023学年秋冬学期		研究生英语			2.0	81	公共学位课								

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



Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/combiomed

MCPNET: Development of an interpretable deep learning model based on multiple conformations of the compound for predicting developmental toxicity

Cheng Cao^{a,b,1}, Hao Wang^{a,1}, Jin-Rong Yang^{a,b}, Qiang Chen^a, Ya-Min Guo^a, Jian-Zhong Chen^{a,*}

^a College of Pharmaceutical Sciences, Zhejiang University, 866 Yuhangtang Rd., Hangzhou, Zhejiang, 310058, China

^b Polytechnic Institute, Zhejiang University, 269 Shixiang Rd, Hangzhou, Zhejiang, 310015, China

ARTICLE INFO

Keywords:

Electrostatic potential

Deep learning

3D multi-conformational surface point cloud

Component visualization

ABSTRACT

The development of deep learning models for predicting toxicological endpoints has shown great promise, but one of the challenges in the field is the accuracy and interpretability of these models. The bioactive conformation of a compound plays a critical role for it to bind in the target. It is a big issue to figure out the bioactive conformation in deep learning without the co-crystal structure or highly precise molecular simulations. In this study, we developed a deep learning framework of Multi-Conformation Point Network (MCPNET) to construct classification and regression models, respectively, based on electrostatic potential distributions on vdW surfaces around multiple conformations of the compound using a dataset of compounds with developmental toxicity in zebrafish embryo. MCPNET applied 3D multi-conformational surface point cloud to extract the molecular features for model training, which may be critical for capturing the structural diversity of compounds. The models achieved an accuracy of 85 % on the classification task and R^2 of 0.66 on the regression task, outperforming traditional machine learning models and other deep learning models. The key feature of our model is its interpretability with the component visualization to identify the factors contributing to the prediction and to understand the compound action mechanism. MCPNET may predict the conformation quietly close to the bioactive conformation of a compound by attention-based multi-conformation pooling mechanism. Our results demonstrated the potential of deep learning based on 3D molecular representations in accurately predicting developmental toxicity. The source code is publicly available at <https://github.com/Superlit-CC/MCPNET>.

1. Introduction

Toxicity is a common drawback of newly designed chemotherapeutic agents. Drug-related toxicities have been the major concerns for the pharmaceutical industry involved in design, synthesis, and development of a new chemical entity. The use of mammalian models for drug discovery and toxicity testing has long been a controversial subject due to the high cost and ethical issues associated with large-scale animal testing. In response to this pressing issue, researchers have sought alternative animal models that are more cost-effective, efficient, and ethically responsible. In recent years, the zebrafish (*Danio rerio*) has emerged as a highly promising animal model for drug discovery and toxicity testing [1–3]. Zebrafish are small vertebrates that are ideal for

large-scale rearing with a short growth cycle and low feeding costs. Stemple et al. identified that 70 % of human genes have at least one obvious zebrafish orthologue [4]. These advantages, together with the recent advances in behavioral testing methods for zebrafish, make this species an attractive option for the identification of toxicity of drugs. Indeed, the zebrafish model provides tremendous advantages for toxicological evaluation of drugs [5–7] to be widely used in the toxicity assessment of human drugs [8–10]. Behavioral tests based on multiple life stages, including adulthood, have been developed over the past 15 years and have the potential to provide large amounts of data for the rapid identification of toxic chemicals. The zebrafish embryo model have been developed to test the developmental toxicity of a compound in environmental assessment and drug development [11].

* Corresponding author.

E-mail address: chjz@zju.edu.cn (J.-Z. Chen).

¹ Joint first authors.

经检索《Web of Science》、《Journal Citation Reports (JCR)》及《中国科学院文献情报中心期刊分区表》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：2024年5月28日)

第1条，共1条

标题:MCPNET: Development of an interpretable deep learning model based on multiple conformations of the compound for predicting developmental toxicity

作者:Cao, C(Cao, Cheng);Wang, H(Wang, Hao);Yang, JR(Yang, Jin-Rong);Chen, Q(Chen, Qiang);Guo, YM(Guo, Ya-Min);Chen, JZ(Chen, Jian- Zhong);

来源出版物:COMPUTERS IN BIOLOGY AND MEDICINE 卷:171 文献号:108037 提前访问日期:FEB 2024

DOI:10.1016/j.combiomed.2024.108037 出版年:MAR 2024

入藏号:WOS:001200434500001

文献类型:Article

地址:

[Cao, Cheng; Wang, Hao; Yang, Jin-Rong; Chen, Qiang; Guo, Ya-Min; Chen, Jian- Zhong] Zhejiang Univ, Coll Pharmaceut Sci, 866 Yuhangtang Rd, Hangzhou 310058, Zhejiang, Peoples R China.

[Cao, Cheng; Yang, Jin-Rong] Zhejiang Univ, Polytech Inst, 269 Shixiang Rd, Hangzhou 310015, Zhejiang, Peoples R China.

通讯作者地址:

Chen, JZ (corresponding author), Zhejiang Univ, Coll Pharmaceut Sci, 866 Yuhangtang Rd, Hangzhou 310058, Zhejiang, Peoples R China.

电子邮件地址:chjz@zju.edu.cn IDS号:NK8H4 ISSN:0010-4825 eISSN:1879-0534

期刊《COMPUTERS IN BIOLOGY AND MEDICINE》2022年的影响因子为7.7，五年影响因子为6.9。

期刊《COMPUTERS IN BIOLOGY AND MEDICINE》2022年的JCR分区情况为:

Edition	JCR® 类别	类别中的排序	JCR 分区
SCIE	BIOLOGY	7/92	Q1
SCIE	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS	18/110	Q1
SCIE	ENGINEERING, BIOMEDICAL	16/96	Q1
SCIE	MATHEMATICAL & COMPUTATIONAL BIOLOGY	4/55	Q1
SCIE	BIOLOGY, MISCELLANEOUS	N/A	N/A

期刊《COMPUTERS IN BIOLOGY AND MEDICINE》2023年升级版的中科院期刊分区情况为:

刊名	COMPUTERS IN BIOLOGY AND MEDICINE		
年份	2023		
ISSN	0010-4825		
	学科	分区	Top 期刊
大类	医学	2	是
小类	BIOLOGY 生物学	1	-
小类	MATHEMATICAL & COMPUTATIONAL BIOLOGY 数学与计算生物学	1	-
小类	COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS 计算机: 跨学科应用	2	-
小类	ENGINEERING, BIOMEDICAL 工程: 生物医学	2	-

注:

1. 期刊影响因子及分区情况最新数据以JCR数据库、《中国科学院文献情报中心期刊分区表》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。





(12) 发明专利申请

(10) 申请公布号 CN 117373559 A

(43) 申请公布日 2024. 01. 09

(21) 申请号 202311200316.0

(22) 申请日 2023.09.18

(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 陈建忠 曹成

(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 赵杭丽

(51) Int. Cl.

G16C 20/30 (2019.01)

G16C 20/70 (2019.01)

G16C 20/80 (2019.01)

G16C 20/50 (2019.01)

G16B 15/30 (2019.01)

权利要求书2页 说明书6页 附图3页

(54) 发明名称

基于分子多构象的点云多示例深度学习毒性预测方法

(57) 摘要

本发明公开了一种基于分子多构象的点云多示例深度学习毒性预测方法,包括:数据预处理,分子的SMILES序列,基于距离几何和ETKDG方法,生成分子的多构象及其三维坐标和电荷,再通过电荷获取分子的范德华表面静电势,范德华表面的每个静电势点作为分子的特征数据,搭建和训练点云数据的多示例深度学习网络模型,预测分子不同构象的毒性,当分子存在毒性时,可视化其构象,并标注出分子可能与特定靶标相结合的重要特征点。本发明通过多示例注意力模型提取分子的不同构象信息,使得预测结果优于现有模型。本发明可解释性和可视化方面优于其它模型,可解决人工智能的黑箱问题,通过模型可视化的结果可以辅助探索化合物的作用机制。