

解放军总医院青年自主创新科学基金项目

科 研 任 务 书

项目名称： 基于人群队列多组学数据建立常见慢性疾病
机器学习多分类预测模型的研究

项目编号： 22QNFC058

项目级别： 青年自主创新科学基金-扶持项目

项目承担单位： 中国人民解放军总医院第一医学中心

项目负责人： 栾哲

研究起止时间： 2023.01 ~ 2024.12

一、研究目的

研究基于机器学习多分类模型，在团队前期已经获得的人群队列组学数据集（宏基因组学、培养组学）基础上进行数据二次分析，建立二类常见慢性疾病（冠心病、糖尿病）风险预测模型并外部验证，有效避免单一疾病预测模型被不相关的疾病混淆而导致错误分类的问题，提高多疾病预测模型准确性和可靠性，具有重要创新价值和现实意义。

二、研究内容及要求

（一）在已有人群队列多组学数据集基础上，借助机器学习多类分类模型（随机森林模型（RF）、K-近邻分析（KNN）、多层感知器（MLP）、支持向量机（SVM）和图卷积神经网络（GCN））建模，分析比较不同多类分类模型预测能力，发掘同常见慢性疾病发生具有显著关联性的新型健康状态评估指标，包括肠菌丰度、菌种单核苷酸多态性、差异代谢通路指标等；

（二）上述新型肠道微生态健康状态评估指标基础上，汇集队列人群的人口学数据和临床检验数据，开发构建二类常见慢性疾病（冠心病、糖尿病）风险预测模型；

（三）借助开放获取的外部公共人群数据库，对预测模型的检验效能进行外部验证，完成对模型预测准确性的现实优化，最终建立基于人群队列多组学数据建立二类常见慢性疾病机器学习多分类预测模型，为后续疾病超前干预奠定基础。

三、研究进度

2023.01~2023.06 在已建成的人群队列工作基础上，将组学数据、人口学信息、临床检验数据进行数据清洗和标准化；

2023.07~2024.06 完成二类常见慢性疾病（冠心病、糖尿病）机器学习多分类预测模型建模和内部验证，比较不同机器学习多类分类模型疾病预测能力，初步建立基于人群队列多组学数据建立疾病机器学习多分类预测模型；

2024.07~2024.12 根据结果对模型进行外部验证并优化。撰写文章，对项目进行总结，完成研究报告及结题评审。

四、成果形式

（一）基于人群队列多组学数据建立二类常见慢性疾病（冠心病、糖尿病）机器学习多分类预测模型（论文或专利形式）；

（二）发掘与疾病发生具有显著因果关联的 5-10 个新型健康状况评估指标（研究报告或论文形式）；

（三）在国内外期刊发表或投稿论文 1-3 篇（论文形式）。

五、项目组成员

序号	姓名	所在单位	职称	专业
1	栾哲	一中心消化内科	主治医师	消化系病
2	孙刚	一中心消化内科	主任医师	消化系病
3	吴俊岭	解放军医学院	在读研究生	消化系病
4	张瀚文	解放军医学院	在读研究生	消化系病
5	陈怡	解放军医学院	在读研究生	消化系病

6	赵志壮	解放军医学院	在读研究生	消化系病
7	张嘉琪	解放军医学院	在读研究生	消化系病

六、研究经费

序号	科目	经费明细 (万元)
1	设备费	0
2	材料费	0
3	外部协作费	4.5
4	燃料动力费	0
5	会议/差旅/国际合作与交流费	1
6	出版/文献/信息传播/知识产权事务费	3
6	劳务费	1.5
7	专家咨询费	0
8	其他支出	0
合计		10

七、其他

本项目的研究成果，及其相关知识产权的权属，归解放军总医院所有。

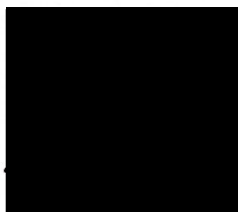
附页

建议密级：绝密□ 机密□ 秘密□ 公开□

主要承担单位意见

承办单位:

承办人:



2023年9月26日

项目分管部门意见

承办单位:

承办人:



20 年 月 日