

华中科技大学创新创业训练计划

项目中期检查表

项目编号	202410487175
项目名称	基于多模态大模型的肺栓塞患者溶栓风险智能诊断
项目负责人	尹荷
所在院（系）	人工智能与自动化学院
实施时间	2024-5-29
联系电话	19839916279
填表时间	2024-11-25

华中科技大学本科生院制

一、项目研究进展情况说明

项目研究的具体时间线如下：

2024. 04~2024. 05: 对项目相关内容进行调研，确定并完成了大创项目的申报，明确了该项目的大致内容、探讨了项目的大体思路与技术路线。学习了 机器学习、神经网络的基础知识以及 python 语言。

2024. 06~2024. 07:学习了相关肺栓塞诊疗方面的医学知识，查询与肺栓塞溶栓治疗出现的不良后果方面的资料，寻找适合作为数据集的相关肺栓塞治疗数据，与同济医学院的老师进行沟通学习。学习了目标检测的基本方法，初步地完成了肺栓塞诊断模型框架的搭建。调研了肺栓塞诊断方法的相关论文。

2024. 08~2024. 09:调研了神经网络相关算法基本思路与方法并进行实践。搭建神经网络并基于肺栓塞和正常人的 ECG 数据进行训练和调整，得到预期的基于 ECG 数据对肺栓塞患者的预测模型。

2024. 10~2024. 11:在预测肺栓塞的基础上着手处理 EHR 数据，探索 XGBoost 多分类模型在肺栓塞溶栓风险方面的预测，搭建基于 EHR 数据对肺栓塞患者溶栓死亡和出血风险的预测。与此同时，讨论下阶段工作计划，完成中期检查报告的撰写。

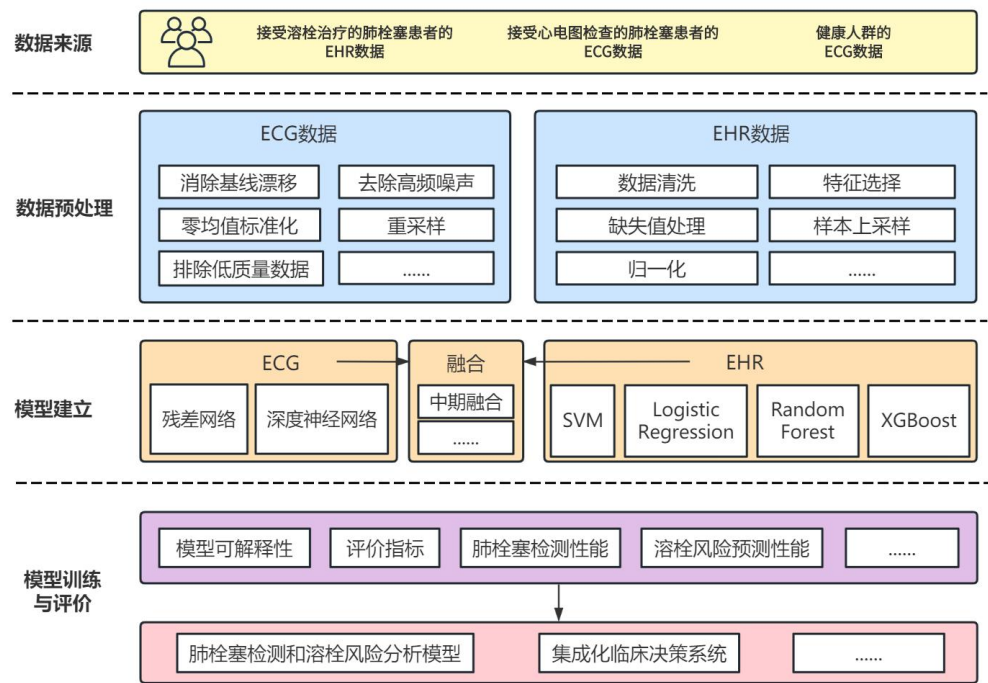


图 1 研究路线框架图

二、已取得的阶段性成果

3.1 ECG 心电图数据预测

肺栓塞预测结果，模型展现出了 67.45% 的准确率。特异性为 0.854，敏感性为 0.419，阳性预测值为 0.668，阴性预测为 0.677，F1 分数为 0.515。模型在测试集上的 AUC 为 0.722。模型的 ROC 曲线以及 AUC 如下。

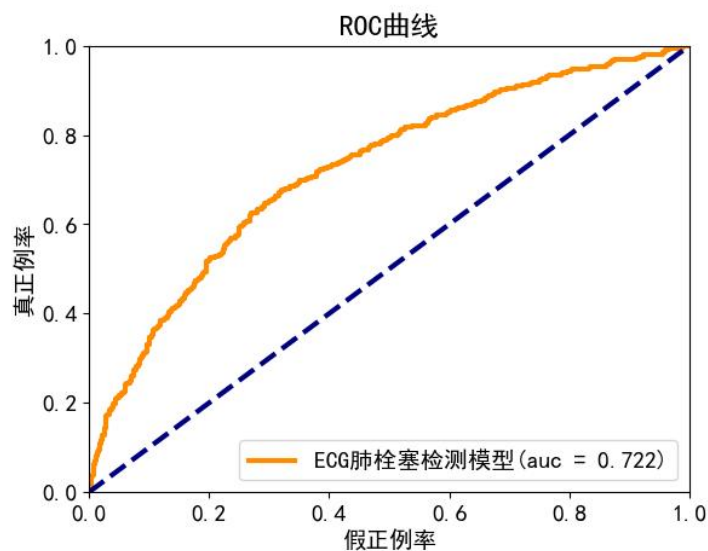


图 3.1 肺栓塞检测模型的 ROC 曲线和 AUC 值

3.2 基于 EHR 数据三分类（正常、死亡、出血）预测

我们对于 EHR 数据，使用 MissForest 算法进行填补后，再将其按照 8: 2 的比例分为训练集和数据集，对训练集通过 smotenc 和 clustercentroids 混合采样得到新的训练集，最后通过使用 xgboost 模型来进行预测。

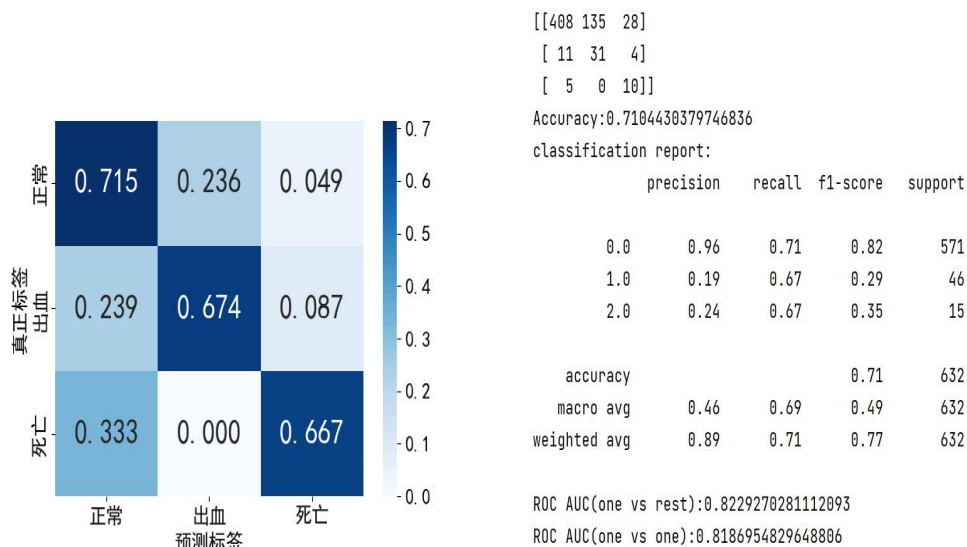


图 3.2.1 三分类预测的混淆矩阵和预测结果

测溶栓后三分类模型的 shap 图特征如下:

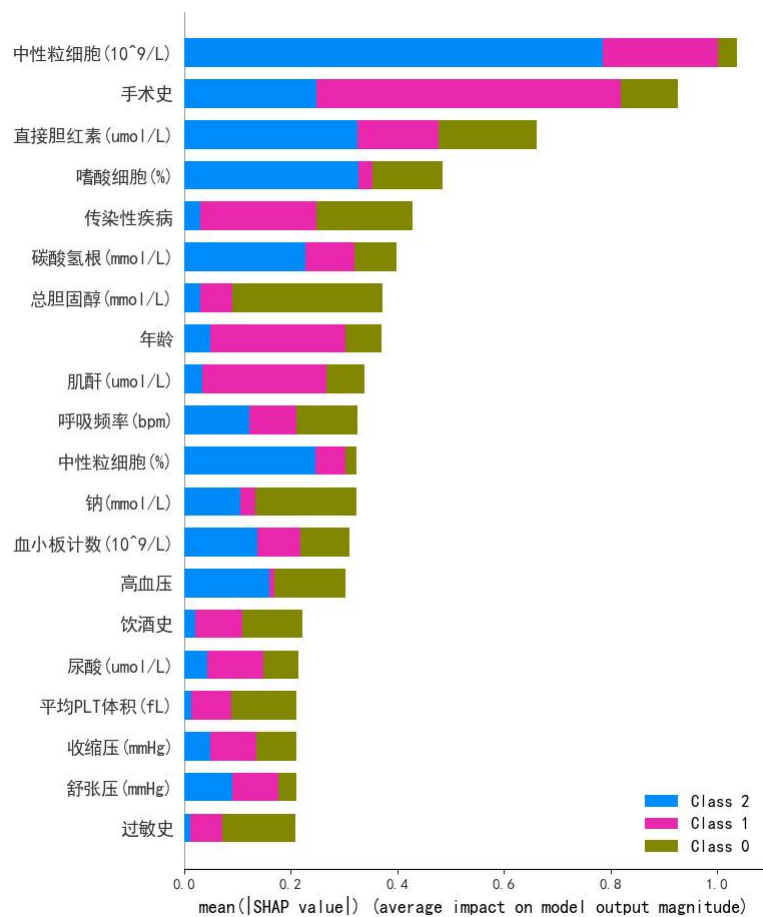


图 3.2.2 三分类 shap 图特征结果

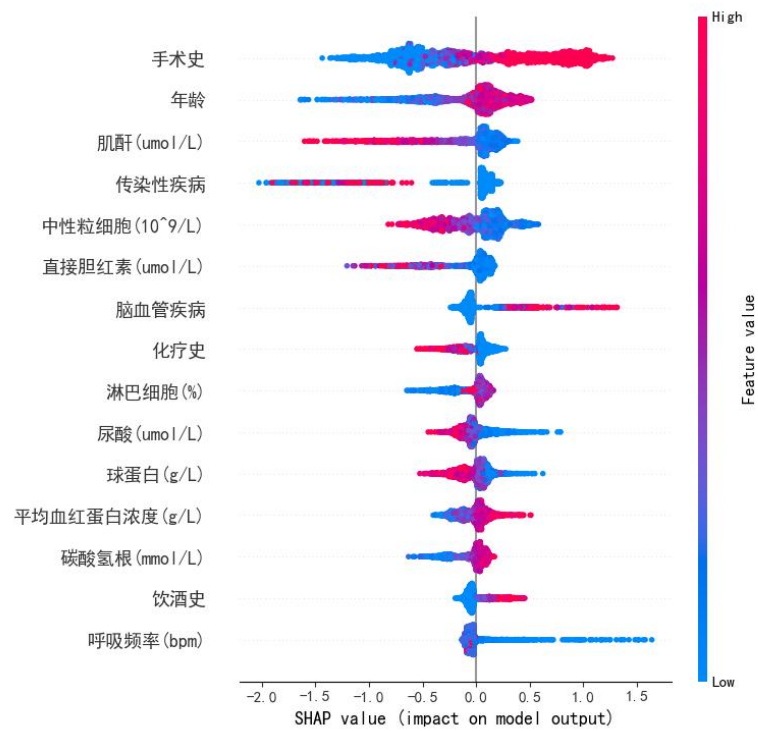


图 3.2.3 出血风险重要性排名前二十特征

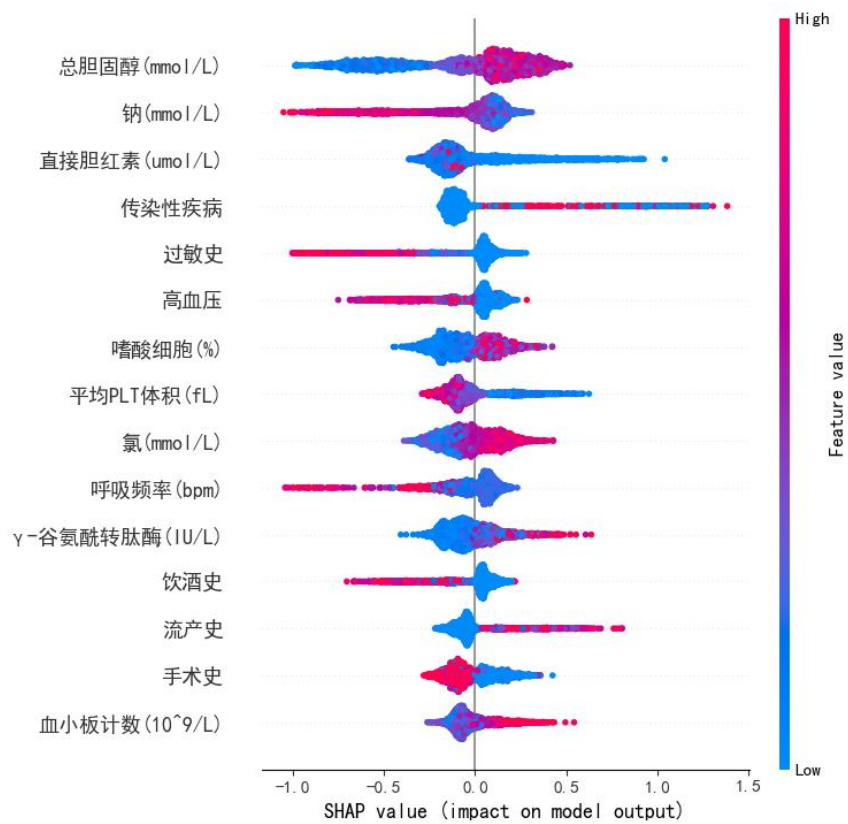


图 3.2.4 死亡风险重要性排名前二十特征

三、目前存在问题

4.1 硬件设施缺乏

我们组的问题主要集中在硬件设施的问题，由于缺乏 GPU 算力资源，有许多很好的深度学习算法无法使用，很多都只能采用传统的机器学习算法，这对我们当初构建模型的实现提出了很大的挑战。

4.2 数据集缺乏

由于是医学相关的研究，可以使用的数据集比较有限且不全面。缺乏数据也一直是困扰我们组的问题。导致训练模型的结果与预期还存在些差距。

4.3 医学知识缺乏

由于是非医学专业，我们组只能通过和同济老师沟通来进行一些模型合理性和准确性的分析，导致模型调参效果不高。另外我们还有做预测溶栓并发症的想法，但是对于并发症判定标准还有待探究。

四、经费使用情况说明

已获资助 经费	10000 元
已使用经 费	5000 元
经费使用 明细	专利费 5000 元

五、下阶段研究计划及主要措施

整体上，我们想要先实现通过 ECG 数据和 EHR 数据对肺栓塞患者的溶栓风险进行融合预测，然后在此基础上进行进一步的研究，构建肺栓塞溶栓诊疗及风险预测综合系统。具体的时间线安排如下：

2024.11~2024.12: 优化肺栓塞对死亡和出血的预测，使其预测结果更加精确、更加贴合医学实际，并在此基础上尝试加入对应患者的 ECG 数据对模型进行改进。完成预期的融合预测模型，并对效果进行检验。

2025.01~2025.03: 在完成基于 ECG 数据和 EHR 数据对肺栓塞患者的溶栓风险进行融合预测的基础上，综合结果，构建肺栓塞溶栓诊疗及风险预测综合系统。

2025.04~2025.05，完善研究内容，优化参数，提高系统预测的稳定性与准确率，分析总结研究结果，准备结题。

项目负责人（签名）

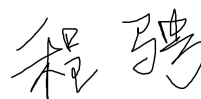


2024 年 11 月 25 日

六、项目指导教师意见

大创团队较好的完成了预期研究目标，目前已完成了 2 篇专利撰写，搭建了 ECG 心电图数据预测模型和基于 EHR 数据三分类（正常、死亡、出血）预测模型，并且提出了通过 ECG 数据和 EHR 数据对肺栓塞患者的溶栓风险进行融合预测，在此基础上进行进一步的研究，构建肺栓塞溶栓诊疗及风险预测综合系统的想法。希望可以进一步形成论文成果。

指导教师（签名）



2024 年 11 月 25 日

七、项目负责人所在院（系）审核意见

通过

审核人（签名） 周纪杰

2024 年 11 月 28 日