

项目编号：

西安交通大学

大学生创新训练项目  
结题报告书

项目名称：基于多元数据分析的重症监护室  
病人健康状态预警方法研究

起止年月：

负责人：

手机：

邮箱：

项目成员：

指导老师：

批准经费： 已用金额：

填报日期：2023 年 6 月 3 日

报告题目:

学生姓名:

指导老师:

中文摘要:

英文摘要:

关键词:

# 目录

1	绪论	1
2	数据集分析及预处理	2
2.1	数据集介绍	2
2.2	数据预处理	2
2.2.1	数据提取	2
2.2.2	数据插补	2
2.2.3	异常值矫正	2
2.2.4	数据填补	2
2.2.5	实时状态数据标记	2
3	模型训练与评估	3
3.1	数据集划分	3
3.2	模型评估	3
4	模型可解释性分析	3
4.1	SHAP 值基本原理	3
4.1.1	加性特征归因方法	3
4.1.2	加性特征归因方法中的唯一解	3
4.1.3	SHAP(SHapley Additive exPlanation) 值	3
4.1.4	LIME 解释方法	3
4.2	基于 SHAP 值的可解释性分析	3
4.2.1	整体数据可解释性分析	3
4.2.2	单个数据可解释性分析	3
5	结论与展望	3
A	附录	5
A.1	预处理后 SIC 和 DIC 数据量	5
A.2	数据集划分结果	5
A.3	模型评估结果	5
A.4	基于 SHAP 值的可解释性分析	5

# 1 绪论

## 2 数据集分析及预处理

### 2.1 数据集介绍

### 2.2 数据预处理

#### 2.2.1 数据提取

数据全称	缩写	MIMIC-IV 数据编号
1. 血小板计数 (Platelet Count)	PLT	227457
2. 凝血酶原时间 (Prothrombin time)	PT	227465
3. 凝血酶原时间的国际标准化比值 (International Normalized Ratio)	INR	227467
4. D-二聚体 (D-Dimer)	D-Dimer	225636
5. 纤维蛋白原 (Fibrinogen)	FIB	227468
6. 二氧化碳分压 (Venous CO2 Pressure)	pCO2	226062 (动脉)
7. 酸碱度 (pH)	pH	223830 (动脉)
8. 氧分压 (Venous O2 Pressure)	pO2	226063 (动脉)

表 1: 相关数据对应缩写及编号

#### 2.2.2 数据插补

#### 2.2.3 异常值矫正

#### 2.2.4 数据填补

#### 2.2.5 实时状态数据标记

### 3 模型训练与评估

#### 3.1 数据集划分

#### 3.2 模型评估

### 4 模型可解释性分析

#### 4.1 SHAP 值基本原理

##### 4.1.1 加性特征归因方法

##### 4.1.2 加性特征归因方法中的唯一解

##### 4.1.3 SHAP(SHapley Additive exPlanation) 值

##### 4.1.4 LIME 解释方法

#### 4.2 基于 SHAP 值的可解释性分析

##### 4.2.1 整体数据可解释性分析

##### 4.2.2 单个数据可解释性分析

### 5 结论与展望

## 参考文献

- [1] Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2023). MIMIC-IV (version 2.2)[DB]. PhysioNet. <https://doi.org/10.13026/6mm1-ek67>.

## A 附录

### A.1 预处理后 SIC 和 DIC 数据量

时间段长度	插补策略	填补策略	SIC 标签数量	DIC 标签数量
4h	均值插补	临近填补	827 (占比 89.89%)	629 (占比 68.37%)
8h	均值插补	临近填补	1885 (占比 87.55%)	1345 (占比 62.47%)
12h	均值插补	临近填补	1823 (占比 86.77%)	1267 (占比 60.30%)

表 2: 预处理后的 SIC 和 DIC 数据量

### A.2 数据集划分结果

### A.3 模型评估结果

### A.4 基于 SHAP 值的可解释性分析