

# 基于多模态融合的救援人员 PTSD 风险评估

## 摘要

救援人员经常参与搜救、维稳、反恐等危险性工作，极易造成创伤，从而诱发创伤后应激障碍（PTSD）。为了避免救援人员发生 PTSD，提早干预，最大可能保障救援人员身心健康。为此，本文基于多模态数据，建立了救援人员 PTSD 的风险评估模型，主要内容如下。

针对问题一，首先，本文针对量表数据进行数据预处理和特征工程，剔除掉了异常数据。其次，绘制了数据分布直方图和频数统计表，初步剔除了信息熵较少的 3 个变量。通过特征提取，得到了  $PCL_3$ 、所属地区、是否少数民族等变量。接下来，为了探究救援人员 PTSD 的影响因素，本文将  $PCL_3$  作为新的响应变量，对变量进行了 Spearman 秩相关性分析，查看变量间的相关性，并建立了 Lasso-Logistic 回归模型。通过调参，确定最优的  $\lambda = 1/6$ ，模型在测试集上的 AUC 值为 0.89，预测效果较好。模型保留了包括年龄、文化程度、家庭人均月收入等在内的 10 个变量，其中年龄、ASD 再体验、ASD 回避、ASD 警觉、ASD 总分为危险因素，其余 5 个变量为保护因素，并进一步进行了交互作用分析。

针对问题二，本文设计了 8 种多模态数据的 PTSD 风险评估模型。在问题一的基础上，首先收集影响救援人员患 PTSD 的显著影响变量，包括分类数据和连续数据，同时对脑电波时间序列数据进行清洗、缺失值插补、滤波去噪，接着从早期融合（SVM、MLP、Logistic），中期融合（MLP+RNN+DNN+Logistic、MLP+LSTM+DNN+Logistic、MLP+GRU+DNN+Logistic），晚期融合（MLP+GRU、MLP+CNN）三个特征融合阶段出发，设计了 8 种救援人员 PTSD 风险评估方案，并比较其性能，结果显示中期融合 MLP+GRU+DNN+Logistic 模型组合对评估救援人员 PTSD 患病风险的效果最佳，进而设计了个性化的评估流程用于早期筛查预防。

综上，本文在多模态数据的基础上构建了个性化救援人员 PTSD 风险评估模型。实现了救援人员 PTSD 的精准预判，对救援人员心理健康保障具有重要的应用价值。

**关键词：**救援人员；PTSD；Logistic 回归；多模态数据；特征融合模型

# 目 录

摘 要	I
一、问题综述	1
1.1 研究背景及意义	1
1.2 研究现状	3
1.3 问题描述	5
二、研究思路	6
2.1 ASD 与 PTSD 的关系	6
2.2 PTSD 影响因素分析	6
2.3 PTSD 风险预测模型	6
三、研究方法	8
3.1 Spearman 秩相关系数分析	8
3.2 Logistic 回归模型	9
3.3 SVM	11
3.4 MLP	11
3.5 LSTM	13
3.6 GRU	14
3.7 CNN	14
四、数据预处理与统计分析	17
4.1 附件 1 数据预处理	17
4.2 附件 1 数据特征提取	23
4.3 脑电波数据预处理	24
4.3.1 查看数据	24
4.3.2 异常值处理	25
4.3.3 数据去噪	27
五、模型的建立与求解	28
5.1 基于 Logistic 回归的救援人员 PTSD 发病相关影响因素分析	28
5.1.1 Spearman 秩相关性分析	28

5.1.2 Lasso-Logistic 回归建模	29
5.1.3 救援人员 PTSD 发病相关影响因素分析	31
5.2 基于多模态数据的 PTSD 风险评估模型	35
5.2.1 数据融合原理	35
5.2.2 早期融合模型	36
5.2.3 中期融合模型	37
5.2.4 晚期融合模型	39
5.2.5 个性化风险评估模型	41
六、结论与建议	43
6.1 结论	43
6.2 建议	44
参考文献	45

---

# 一、问题综述

## 1.1 研究背景及意义

### （一）研究背景

在自然灾害和意外安全事故发生，或者面对社会安全挑战时，救援人员的身影总是随处可见——包括消防员、警察、军事人员、医护急救人员和救援志愿者等。他们在第一线参与搜索与营救，冒着生命危险，奋勇前行。救援人员不仅拯救了人们的性命，保护了人们的家庭，还在打击和阻止恐怖主义活动、维护国家安全中发挥着关键作用。他们的存在为我们的社会带来了极大的稳定与秩序，为人们的安全保驾护航，彰显了人道主义精神的光辉。

然而，正因为救援人员常常处在第一线，他们将直面最原始的意外和灾难现场：血淋淋的受害者、惨烈的场景、最恐惧的情形，这一切都可能在他们的心理留下不可磨灭的创伤，引发严重的心理疾病和精神疾病，进而影响他们的日常工作和生活。救援人员的工作环境和职业压力非常大，经常需要长时间高强度工作，并且在处理紧急情况时面临高风险暴露。例如，消防员在扑灭火灾时面临着随时可能发生的爆炸和建筑物倒塌的危险，医护急救人员在处理突发疾病和事故现场时面对的是生命垂危的患者，这些都给他们带来了巨大的心理压力。因此，救援人员因为其特殊的职业内容容易出现应激相关心理障碍、焦虑与抑郁等情绪问题以及职业倦怠，其中，创伤后应激障碍（Post-traumatic Stress Disorder, PTSD）是救援人员最常见的心理问题<sup>[1]</sup>。

PTSD 是指个体经历、目睹或遭受一个或多个自身或他人有关的死亡、受到死亡威胁、严重创伤等延迟出现或持续存在的精神障碍<sup>[2]</sup>。PTSD 的症状表现多样，包括重复体验（即回忆或梦到创伤经历）、对创伤伴有的刺激持续回避、警觉性增高以及对周围的认知抱有负面情绪等。值得说明的是，这些症状都是人类正常的生理心理应激反应，但是若这些症状长期存在，并对个体正常的生活、工作和学习造成了破坏性的影响，才会被诊断为 PTSD。患有 PTSD 的人往往同时会患有焦虑、抑郁以及其他精神障碍类疾病，长期患有 PTSD 可能会导致死亡。大量研究表明，救援人员的 PTSD 患病率远高于其他群体，不同种类救援人员与普通人群的 PTSD 患病率比较结果如图 1 所示。普通人群患 PTSD 的概率大约为 6.8%<sup>[3]</sup>，而消防人员的患病率为 14%<sup>[4]</sup>，警察的患病率为 13%<sup>[5]</sup>，军事人员的患病率为 12%<sup>[6]</sup>，在新冠肺炎期间的一项调查发现医护人员 PTSD 的患病率为 10%<sup>[7]</sup>。

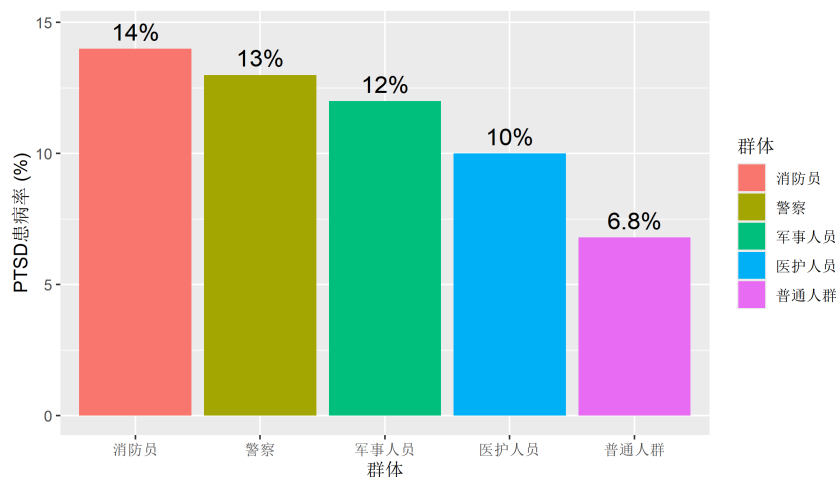


图 1.1 不同救援人员与普通人群的 PTSD 患病率比较

受调查地域、文化、样本容量、调查方式等的影响，不同研究的具体 PTSD 患病率有所差别，但都表明救援人员 PTSD 的患病率要远高于普通人群的患病率。

如果救援人员患有 PTSD，一方面，该病会影响他们的日常生活、家庭生活和身心健康，另一方面会影响他们的救援工作，影响救援效率，甚至导致自杀，造成高技能、高素质救援人才的流失，对国家和人民是不可估量的损失。因此，若在创伤事件发生后能通过一些心理评定工具来初步评定个体的心理健康状况，将有助于筛选出 PTSD 高危人群，从而有针对性地对高危人群提供有效的干预策略。

新冠肺炎疫情后，我们越来越重视医护工作者的心理健康，也有越来越多的地区强调消防员心理服务体系的规范化建设，救援人员的心理健康问题正逐渐得到重视。对救援人员 PTSD 的影响因素进行分析研究，明确影响因素及其作用机制；同时，构建救援人员 PTSD 风险预测模型，获得良好的风险识别能力，通过这些以期能够提前筛选出重点救援人员，提供有效的 PTSD 干预手段，有效预防 PTSD 的发生，也有助于健全救援人员心理援助体系，为救援人员的心理健康提供保障。

## （二）研究意义

### I 理论意义

目前，关于 PTSD 的影响因素的研究主要集中于医疗领域，围绕临床上因创伤患病的患者进行了充分的研究，也存在部分关于急诊医护人员 PTSD 方面的研究，但整体上关于救援人员 PTSD 的研究仍较少，关于 PTSD 风险预测的研究更是非常缺乏。故本文对救援人员 PTSD 影响因素及风险预测的研究能够补充该领

---

域的研究资料，丰富研究结果，为后续的研究提供借鉴。

本文的研究对象是救援人员这个大范畴，涵盖医护人员、警察、军事人员等，与以往常常聚焦于某一职业人员或者某一地区人员不同，是从整体上对这类人员患 PTSD 的相关研究，故本研究具有归纳、概括的作用，在一定程度上为后续范围内的研究提供理论参考。

本文在构建救援人员 PTSD 风险预测模型时，结合了多种不同的数据类型，即基于多模态数据的方法，在该领域用此方法开展的研究还很缺乏。在此过程中，我们整合了心理测评数据和生理监测数据。在数据融合的基础上，本文探索运用先进的学习算法，构建一个高精度的 PTSD 风险预测模型。

## II 现实意义

本文基于统计数据，重点分析救援人员 PTSD 的保护因素和危害因素，明确因素增加和降低救援人员 PTSD 的作用机制，这对于救援人员 PTSD 的干预方向具有重要指导作用，为救援人员 PTSD 干预方案的制定提供了科学依据和理论支持，有助于调节救援人员遭遇创伤后的心理状态，有助于帮助患有 PTSD 早期症状的救援人员，避免他们的精神健康持续恶化。

本文在原有统计数据的基础上，进一步结合脑电波数据，构建了救援人员 PTSD 风险预测模型。该模型旨在实现救援人员 PTSD 的精准预判，用于早期筛查和干预，从而有助于 PTSD 的早期发现和干预。这不仅能提升救援人员的心理健康水平，还对国家相关救援人员心理服务体系的规范化建设具有重要指导意义。本文构建的 PTSD 风险预测模型不仅具有较强的风险识别能力，还为救援人员 PTSD 的干预提供了有效的证据支撑，具有重要的应用价值。模型结合了临床脑电波数据，增加了分析结果的科学性和真实性，为早期 PTSD 筛查提供了合理依据。这种结合多种数据源的方法，使得模型在预测 PTSD 风险时更加准确和全面。

通过本文对救援人员 PTSD 影响因素的分析及建立 PTSD 风险预测模型，我们能够实现对高风险救援人员的早期鉴别与有效干预，从而更好地保障和维护救援人员的心理健康，提高他们在执行任务时的效率和安全性。

## 1.2 研究现状

关于 PTSD 影响因素方面的研究，研究对象主要围绕遭遇意外事故或者经历重大手术的临床创伤患者，关于救援人员的研究多数集中于一线的医护工作者，存

---

在少量关于消防员的研究。其中, logistic 回归分析和分层回归分析是常用的多因素分析方法。

李艳芳、钟仲英等<sup>[8]</sup>以江门市新会区人民医院复诊的严重创伤患者为研究对象,运用 Logistic 回归分析的方法分析严重创伤患者 PTSD 的影响因素,发现高家庭收入水平、长创伤时间和高积极心理资本得分是其保护因素。石岩、杨罗东等<sup>[9]</sup>采用二元 Logistic 回归分析的方法对经历亲人亡故 PTSD 患者的影响因素进行研究,结果表明吸烟和饮酒是这类 PTSD 患者的保护因素。罗娟、伍娟等<sup>[10]</sup>对主动脉夹层术后患者发生创伤后应激障碍的现状和影响因素进行探讨,Logistic 回归分析表明家庭人均月收入、患者的心理韧性和社会支持对术后患者的心理状态有重要影响。周守志、尹弘青等<sup>[11]</sup>采用分层回归分析对三级综合性医院急诊护士 PTSD 的影响因素进行研究,结果显示是否是省级专科护士、生理领域、心理领域、社会领域、环境领域、睡眠质量是影响急诊科护士 PTSD 的重要因素。朱白鹭、刘欢等<sup>[12]</sup>采用分层回归分析法对新型冠状病毒肺炎定点收治医院的医护人员 PTSD、心理弹性和生活质量的关系进行研究,发现 PTSD 会降低生活质量,心理弹性能够保证生活质量,建议在突发公共卫生事件中应通过提升医护人员的心理弹性、预防和治疗 PTSD 提升其生活质量。孔嘉文、江俊颖等<sup>[13]</sup>对上海市消防人员 PTSD 的影响因素进行研究,分层回归分析的结果表明消防救援人员的易感性因素和社会心理因素会影响 PTSD 症状的发生,一般自我效能感、职业认同、朋友支持、集体自尊是保护性因素。Yijin Wu、Zhenwei Dai 等<sup>[14]</sup>应用单因素分析和二元 logistic 回归分析研究新冠肺炎患者激增期间中国大陆医护人员患 PTSD 的心理社会因素,结果表明专注度、心理弹性和感知到的社会支持对于处于疫情一线医护人员的健康心理建设有重大作用。裘懿琳<sup>[15]</sup>应用线性回归分析新冠肺炎疫情下中高风险地区居民创伤后 PTSD 的影响因素,发现居民年龄和消极应对态度能够显著促进 PTSD,而教育水平、月收入以及被隔离情况可以抑制 PTSD。Xiangyuan Chu、Xiu Dai 等应用随机效应模型分析、质量评估、亚组分析和敏感性分析等方法<sup>[16]</sup>探究病重儿童父母 PTSD 的患病率和危险因素趋势。

关于 PTSD 风险预测方面的研究目前仍然相对较少,尚未形成广泛应用的标准化方法,但应用机器学习的方法正逐渐形成一种趋势,该领域存在着巨大的研究空间和发展潜力。

黄欢、杨佳佳等<sup>[17]</sup>以脑卒中患者为研究对象,考虑文化程度、内向人格、ADL 重度障碍、吞咽功能障碍等影响因素,构建出脑卒中患者发生 PTSD 的风险预测列线图,直观地展示该模型具有良好的一致性和区分度,可以辅助临床医护人员筛

---

选 PTSD 高风险的脑卒中患者。张津、刘川源等<sup>[18]</sup>分析外伤性骨折并发 PTSD 的影响因素，根据多因素分析结果，选取社会支持评分、损伤严重程度、内外向评分和院内感染四个因素建立线性预测模型，用受试者操作特征（ROC）曲线分析预测价值，分析结果表明建立的预测模型对 PTSD 的判断具有较高准确性。预测价值邓傲骞、杨燕贻等<sup>[19]</sup>用机器学习算法构建长沙消防员患 PTSD 的预测模型，考虑包含性别、学历、坚韧性和反思深省等在内的 11 个特征，通过 5 折交叉验证比较 5 中分类算法的预测能力，最终选取随机森林方法构建 PTSD 预测模型，模型的平均精确率为 90.1%。Joyce Zhang、Saber Sami 等<sup>[20]</sup>基于随机森林提出了一个创伤后 PTSD 半年到一年的预测模型，并用 Shapley 值和偏依赖图（PDP）来对模型结果进行解释分析。Adam Horwitz、Kaitlyn McCarthy 等<sup>[21]</sup>基于分类回归树（CART）确定了一组最小的创伤相关症状，并证明通过在创伤后一周三次评估这些症状的方式可以筛选出有患 PTSD 风险的人。

### 1.3 问题描述

第一问，根据企业提供的 28 种特征数据以及一种因变量数据，分析救援人员 PTSD 的影响因素及其影响机制，包括保护因素和危险因素，即能够增加和抑制救援人员 PTSD 的因素及其作用机制，同时尝试挖掘保护因素和危险因素的交互作用。

第二问，在已有数据的基础上，结合新增的脑电波监测数据，即在多模态数据融合的基础上，构建合理、有效的 PTSD 风险预测模型。需要提供模型有效的性能证据，证明模型能够实现 PTSD 的风险判别，并解释运用模型对救援人员 PTSD 进行筛查的机制。



---

## 二、研究思路

### 2.1 ASD 与 PTSD 的关系

ASD 的全称是急性应激障碍, PTSD 的全称是创伤后应激障碍。ASD 和 PTSD 都是由创伤经历引发的心理障碍, 但它们在诊断标准、持续时间和症状表现上有所不同。

ASD 是一种在创伤事件发生后短期内出现的心理障碍。它通常会在事件发生后 3 天到 4 周之间表现出来。一般而言, ASD 总分是由 ASD 分离、再体验、回避、警觉等四个变量通过直接求和或某种方式求和得到的。

PTSD 是一种在创伤事件发生后较长时间内持续存在的心理障碍。其症状通常在事件发生后的 1 个月或更久才会显现, 并且症状持续时间超过 4 周。

简单来说, ASD 和 PTSD 都是由创伤经历引发的心理障碍, 但 ASD 是较为短期的反应, 而 PTSD 是长期持续的状态。

### 2.2 PTSD 影响因素分析

首先对附件 1 中的数据进行预处理, 包括数据清洗、缺失值处理以及异常值处理等得到合理正规的数据, 然后通过描述性分析以及数据可视化等操作对数据做初步分析与展示, 并考虑是否需要部分变量数据做一定处理, 例如特征提取和数据转换, 以期适用后续的模型方法。

为了更好的分析 PTSD 的影响因素, 考虑先用相关性分析方法对 PCL 与自变量之间的相关性进行分析, 对会影响 PCL 的特征做一个初步了解。然后, 在以 PCL 为因变量, 特征为自变量应用 Logistic 回归模型进行分析, 考虑到本文研究的因素数量多, 单纯用 Logistic 回归模型可能会产生过拟合, 故考虑加入惩罚项, 提高泛模型的泛化能力。最后, 综合相关性分析和模型回归的结果, 分析 PTSD 的保护因素、危险因素以及因素的交互作用。

### 2.3 PTSD 风险预测模型

我们需要在第一问的基础上, 加入新数据——脑电波数据, 在数据融合的基础上构建 PTSD 风险预测模型。同样地, 首先需要对脑电波数据进行数据处理, 除了异常值、缺失值处理等常规操作外, 对于脑电波这种数据还需要考虑数据去噪处理, 得到干净无扰动的数据。

---

结合研究现状，考虑利用机器学习和深度学习模型来构建 PTSD 风险预测模型。同时，考虑结合不同数据融合的方法，包括早期融合、中期融合和晚期融合。故提出多种多模态数据融合模型组合。通过比较不同的组合模型的性能，来确定最佳的 PTSD 风险预测模型。最后基于该模型，解释 PTSD 高风险救援人员的筛选机制。

---

## 三、研究方法

### 3.1 Spearman 秩相关系数分析

斯皮尔曼秩相关系数 (Spearman's rank correlation coefficient), 简称 Spearman 相关系数, 是秩相关 (rank correlation) 的一种非参数度量 (nonparametric measure)。得名于英国统计学家 Charles Spearman, 常用希腊字母  $\rho$  表示。

Spearman 相关系数可以看作是皮尔逊相关系数的非参数版本 (nonparametric version), 皮尔逊相关系数描述的是两个变量之间的线性关系强度, 而 Spearman 相关系数衡量的是两者单调关系的强度, 或者说是利用单调函数来考察两个变量的相关性。计算皮尔逊相关系数时使用的是样本数据值本身, 但计算 Spearman 相关系数使用的是样本数据的排位位次值, 或者说是等级值, 这也是称为秩相关的原因, 所以如果原数据本身不是位次值, 在计算 Spearman 相关系数前需要先计算位次值。另外, 计算 Spearman 相关系数不需要除了数据本身以外的信息, 即不需要先验知识, 所以它是非参数的。

考虑两个随机变量  $X$  和  $Y$ , 如果秩相关系数为正, 则说明  $Y$  随着  $X$  的增加而增加; 如果秩相关系数为负, 则说明  $Y$  随着  $X$  的增加而减小; 如果秩相关系数为 0, 则表示  $Y$  的增减变化跟  $X$  的增减变化没啥关系。当  $Y$  和  $X$  越来越接近严格单调的函数关系时, 秩相关系数在数值上就越来越大。当秩相关系数为 1 或者 -1 时, 就表明  $Y$  随着  $X$  的增加而严格单调增加或单调减小。

通常考察变量间的相关关系时会优先考虑使用皮尔逊相关, 但是如果计算皮尔逊相关系数需要的前提假设不满足, 此时就可以考虑使用 Spearman 相关。比如以下情况:

- 1、数据不服从正态分布, 或者数据呈现非线性关系。
- 2、至少有一方数据是定序类型 (ordinal) 而非数值类型。例如年级变量, 赋值是“一、二、三、...”定序变量值可以排序, 但不能加减。
- 3、数据中存在明显的异常值 (outliers)。与皮尔逊相关不同, Spearman 相关对于异常值不太敏感, 因为它基于排序位次进行计算, 实际数值之间的差异大小对于计算结果没有直接影响。

Spearman 相关系数计算公式如下:

$$\rho = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)} \quad (3.1)$$

其中,  $d_i$  表示第  $i$  个数据对的位次值之差,  $N$  是总的观测样本数。  
或者也可以用如下公式:

$$\rho = \frac{\sum_{i=1}^n (R(x_i) - \overline{R(x)})(R(y_i) - \overline{R(y)})}{\sqrt{\sum_{i=1}^n (R(x_i) - \overline{R(x)})^2 \sum_{i=1}^n (R(y_i) - \overline{R(y)})^2}} \quad (3.2)$$

其中,  $R(x_i)$  和  $R(y_i)$  分别是样本  $x_i$  和  $y_i$  的位次,  $R(x)$  和  $R(y)$  分别表示  $x$  和  $y$  样本的平均位次。

当我们的变量是以变量组的形式存在, 变量组与变量组间 Spearman 相关系数的计算是指内部变量的两两之间的 Spearman 相关系数计算, 为了能直观地观察到变量间的相关关系, Spearman 相关系数热力图是一种常用的可视化工具, 其使用颜色来表示相关系数的大小, 颜色越深表示正相关的程度越大, 颜色越浅表示负相关的程度越大, 无相关关系的变量则使用中性颜色表示。

### 3.2 Logistic 回归模型

Logistic 回归模型作为一种广义线性回归模型, 是一种常用的分类模型。其根据现有数据对分类边界建立回归公式, 以此进行分类。根据类别的个数, 可以使用二分类 Logistic 回归或者多分类 Logistic 回归。其中, 二分类 Logistic 回归最常见。Logistic 回归是以线性回归为理论支持的, 但 Logistic 回归通过 Sigmoid 函数引入了非线性因素, Sigmoid 函数可以将线性模型的结果压缩的  $[0,1]$  之间, 使其拥有概率意义, 从而可以进行比较, 找出最大的可能结果。即单调可微的 Sigmoid 函数作为联系函数, 将线性回归模型的连续预测值与离散的真实标签值联系了起来, 表达公式如式 (3.3) 所示。

$$y = \frac{1}{1 + e^{-(w^T x + b)}} \quad (3.3)$$

其中, Sigmoid 函数表达式如式 (3.4) 所示。

$$y = \frac{1}{1 + e^{-z}} \quad (3.4)$$

整理公式 (3.3) 进一步得到式 (3.5)。

$$\ln \frac{y}{1-y} = w^T x + b \quad (3.5)$$

若将  $y$  视为正例后验概率估计  $p(y = 1|x)$ ，则  $1 - y$  就是其反例后验概率估计  $p(y = 0|x)$ ，则式 (3.5) 可以重写为

$$\ln \frac{p(y = 1|x)}{p(y = 0|x)} = w^T x + b. \quad (3.6)$$

进一步显然有

$$\begin{aligned} p(y = 1|x) &= \frac{e^{w^T x + b}}{1 + e^{w^T x + b}}, \\ p(y = 0|x) &= \frac{1}{1 + e^{w^T x + b}}. \end{aligned} \quad (3.7)$$

其中， $x \in \mathbf{R}^n$  是输入， $y \in \{0, 1\}$  是输出， $w \in \mathbf{R}^n$  和  $b \in \mathbf{R}$  是待求的参数， $w$  称为权值向量， $b$  称为偏置。如式和式所示的条件概率就是我们的二项 Logistic 回归模型。对与给定的输入实例  $x$ ，按照式 (3.7) 可以求得  $p(y = 1|x)$  和  $p(y = 0|x)$ ，再比较这两个条件概率值的大小，将实例  $x$  分到概率值较大的那一类。

在应用中，关键步骤是根据训练数据，通过监督学习的方法估计出 Logistic 回归模型的参数  $w$  和  $b$ 。通过极大似然推导出的损失函数如式 (3.8) 所示。

$$J(\theta) = \sum_{i=1}^m [\mathbf{y}^{(i)} \log(h_{\theta}(\mathbf{x}^{(i)})) + (1 - \mathbf{y}^{(i)}) \log(1 - h_{\theta}(\mathbf{x}^{(i)}))] \quad (3.8)$$

其中， $h_{\theta}(\cdot)$  是 Sigmoid 函数。接下来通过优化算法求解该负对数似然函数的最小值即可得到参数  $w$  和  $b$ 。

通过最小化损失函数得到 Logistic 回归模型的参数后，我们还需要根据测试数据对我们的模型进行评价。分类模型常用的评价指标有准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、F1 值、ROC 曲线及 AUC 值。准确率 (Accuracy) 的计算公式如下：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3.9)$$

其中，TP 表示，TN 表示，FP 表示，FN 表示（下同）。

精确率 (Precision) 的计算公式如下：

$$Precision = \frac{TP}{TP + FP}, \quad (3.10)$$

召回率 (Recall) 的计算公式如下：

$$Recall = \frac{TP}{TP + FN}, \quad (3.11)$$

F1 值的计算公式如下：

$$F1_{score} = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{1}{\frac{1}{2}(\frac{1}{Precision} + \frac{1}{Recall})}. \quad (3.12)$$

---

### 3.3 SVM

支持向量机 (Support Vector Machine, SVM) 是一种常用于二分类任务的强大的监督学习模型。SVM 的核心思想是通过一个非线性的转换 (核函数), 将不可分的原始数据映射到一个高维的可分的特征空间, 在该特征空间中找到一个最佳超平面将数据集中的不同类别分开。这个超平面不仅要能正确分类训练数据, 还要尽量最大化到各类别最近数据点 (称为支持向量) 的距离, 从而提高模型的泛化能力。对于线性不可分的数据, SVM 通过核函数将数据映射到更高维的空间, 使得在这个新空间中数据可以被线性分开。常用的核函数包括:

- 线性核函数:  $K(x_i, x_j) = x_i \cdot x_j$
- 多项式核函数:  $K(x_i, x_j) = (x_i \cdot x_j + r)^d$
- 高斯径向基核函数:  $K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2)$

假设我们有一个训练数据集  $(x_i, y_i)_{i=1}^n$ , 其中  $x_i \in \mathbb{R}^d$  表示第  $i$  个样本的特征向量,  $y_i \in -1, 1$  表示第  $i$  个样本的类别标签。SVM 的目标是找到一个超平面  $w \cdot x + b = 0$ , 使得:

- 对于所有  $y_i = 1$  的样本  $x_i$ , 满足  $w \cdot x_i + b \geq 1$
- 对于所有  $y_i = -1$  的样本  $x_i$ , 满足  $w \cdot x_i + b \leq -1$

换句话说, 我们希望找到参数  $w$  和  $b$ , 使得对所有样本  $(x_i, y_i)$ , 满足

$$y_i(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \dots, n. \quad (3.13)$$

同时, 我们希望最大化超平面到支持向量的距离, 即最大化  $\frac{2}{\|w\|}$ , 相当于最小化  $\frac{\|w\|^2}{2}$ 。这是一个带约束条件的凸二次规划问题, 常用序列最小化优化 (Sequential Minimal Optimization, SMO) 算法进行求解。

### 3.4 MLP

多层感知机 (Multilayer Perceptron, MLP) 是人工神经网络的一种, 是最早期和最简单的深度学习模型之一。它由至少三层节点组成: 输入层、一个或多个隐藏层以及输出层。如图 3.1 所示, 展示了一个简单 MLP 的结构。具体来说:

- 
- 输入层 (Input Layer): 输入层节点数与输入数据的特征数相同。每个节点对应一个输入特征, 负责接收输入信号。
  - 隐藏层 (Hidden Layer): 隐藏层包含若干节点, 这些节点通过权重和偏置进行连接。隐藏层可以有多个, 每层的节点数可以不同。隐藏层的节点使用非线性激活函数 (如 ReLU, Sigmoid 等) 来引入非线性, 使得网络能够学习复杂的模式。
  - 输出层 (Output Layer): 输出层节点数与任务相关。例如, 二分类问题有一个输出节点, 多分类问题有多个输出节点 (对应每个类别一个节点)。输出层通常使用适当的激活函数 (如 Sigmoid 用于二分类, Softmax 用于多分类) 来生成最终的输出。

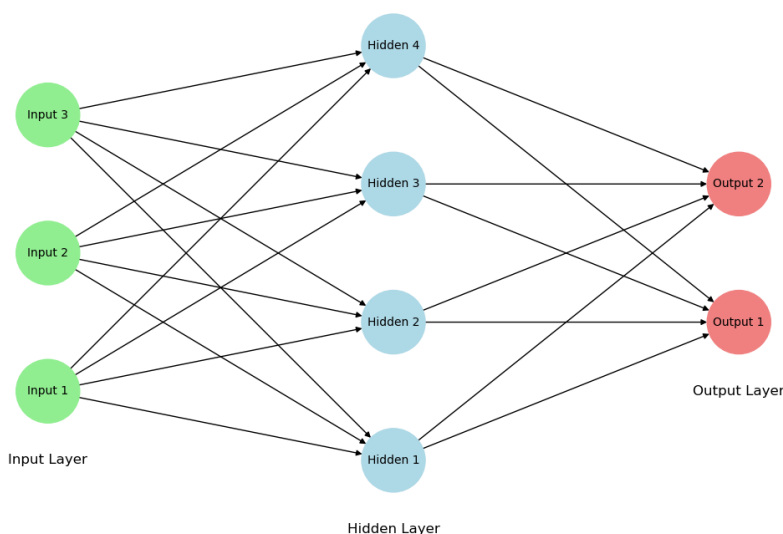


图 3.1 MLP 的结构示例

MLP 通过前向传播 (forward propagation) 和反向传播 (backward propagation) 来学习和优化权重。MLP 结构简单, 便于理解和实现。因为使用了非激活函数所以能够处理非线性数据。但是它往往含有大量的训练参数, 计算量大, 且容易过拟合。对于一些复杂的任务, 其性能可能不如更复杂的深度学习模型。尽管如此, 它的简单性和有效性使得它仍然是许多应用和研究中的首选模型。

### 3.5 LSTM

长短期记忆网络 (Long Short-Term Memory, LSTM) 是一种特殊的循环神经网络 (RNN)。用于解决一般 RNN 中普遍存在的长期依赖问题, 使用 LSTM 可以有有效的传递和表达长时间序列中的信息并且不会导致长时间前的有用信息被忽略 (遗忘)。与此同时, LSTM 还可以解决 RNN 中的梯度消失和梯度爆炸问题。

LSTM 的基本单元称为 LSTM 单元或 LSTM 块, 如图 3.2 所示。每个单元包含三个主要的门: 遗忘门、输入门和输出门。这些门通过控制信息的流动来决定哪些信息应该记住, 哪些应该忘记。其中:

- 遗忘门 (Forget Gate): 决定当前时刻要遗忘多少先前的状态。
- 输入门 (Input Gate): 决定当前时刻要添加多少新的信息到单元状态。
- 输出门 (Output Gate): 决定当前时刻的单元状态要输出多少到隐藏状态。
- 单元状态 (Cell State), 它像传送带一样沿着时间步骤传递信息, 受到各个门的控制进行更新和维护。

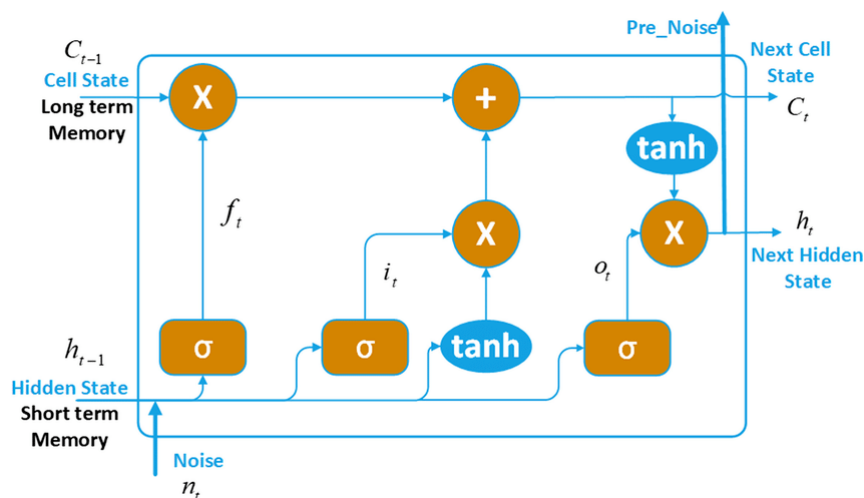


图 3.2 LSTM 的结构示例

通过遗忘门、输入门和输出门, LSTM 能够有效地保留或舍弃信息, 避免梯度消失问题。相比传统的 RNN 模型, LSTM 更能捕捉长时间依赖关系, 能够处理更长的序列数据。LSTM 适用于各种序列数据任务, 如时间序列预测、自然语言处理等。



### 3.6 GRU

门控循环单元 (Gated Recurrent Unit, GRU) 是一种改进的循环神经网络 (RNN)，通过引入门控机制来解决传统 RNN 在处理长序列数据时的梯度消失和梯度爆炸问题。GRU 的结构如图 3.3 所示，包括重置门、更新门和候选隐藏状态。其中：

- 重置门 (Reset Gate)：控制如何组合新输入和先前的记忆。如果重置门接近 0，网络将忽略以前的状态，只使用当前输入。
- 更新门 (Update Gate)：控制从先前状态到当前状态的记忆量。如果更新门接近 1，网络将保留大部分以前的状态。
- 候选隐藏状态 (Candidate Hidden State)：结合当前输入和先前状态，通过重置门调节，以生成新的候选隐藏状态。

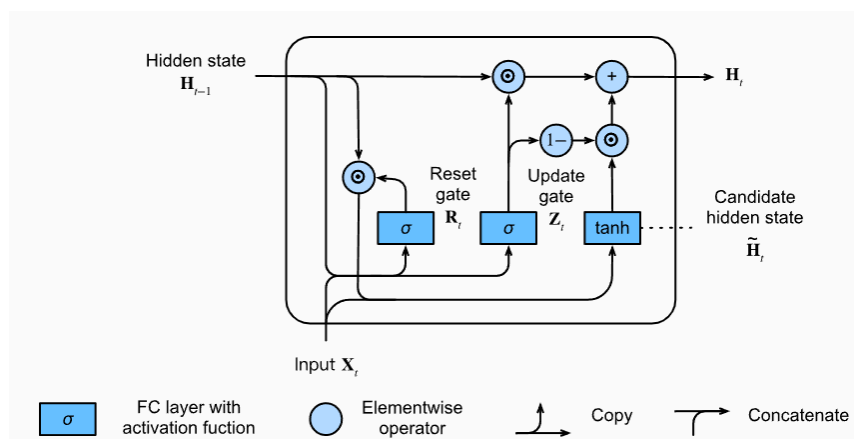


图 3.3 GRU 的结构示例

GRU 被广泛应用于各种序列数据处理任务。相比长短期记忆网络 (LSTM)，GRU 有较少的参数，因为它没有独立的输出门，这使得 GRU 在某些任务上训练速度更快，计算成本更低，性能更好。通过重置门和更新门，GRU 能够灵活地选择性记忆和遗忘信息，更好的捕捉长期依赖。

### 3.7 CNN

卷积神经网络 (Convolutional Neural Network, CNN) 是一种深度学习模型，尤其擅长处理具有网格结构的数据，如图像和视频。CNN 通过卷积层、池化层和

---

全连接层的组合，自动提取数据特征，减少参数量，从而提高计算效率和模型性能。

CNN 的主要组成部分包括卷积层、激活函数、池化层和全连接层，图 (3.4) 展示了 CNN 的基本结构。具体来说：

- 卷积层 (Convolutional Layer)：卷积核 (filter) 在输入数据上滑动，进行点积运算，从而提取局部特征。每个卷积核都有特定的权重，这些权重在训练过程中进行优化。不同的卷积核可以提取不同的特征，如边缘、纹理等。通过卷积操作将输出一个特征图 (feature map)。
- ReLU (Rectified Linear Unit) 激活函数：是 CNN 中最长用的激活函数，计算简单，能有效解决梯度消失问题。
- 池化层 (Pooling Layer)：最大池化 (Max Pooling) 和平均池化 (Average Pooling) 是常用的两种。通过池化可以降低数据维度，减少计算量，在一定程度上防止过拟合。
- 全连接层 (Fully Connected Layer)：每个神经元与上一层的所有神经元相连。通常位于网络的最后几层，用于将卷积层和池化层提取的特征映射到输出类别。
- 输出层：使用适当的激活函数（如 Softmax）生成最终的分类结果或回归结果。

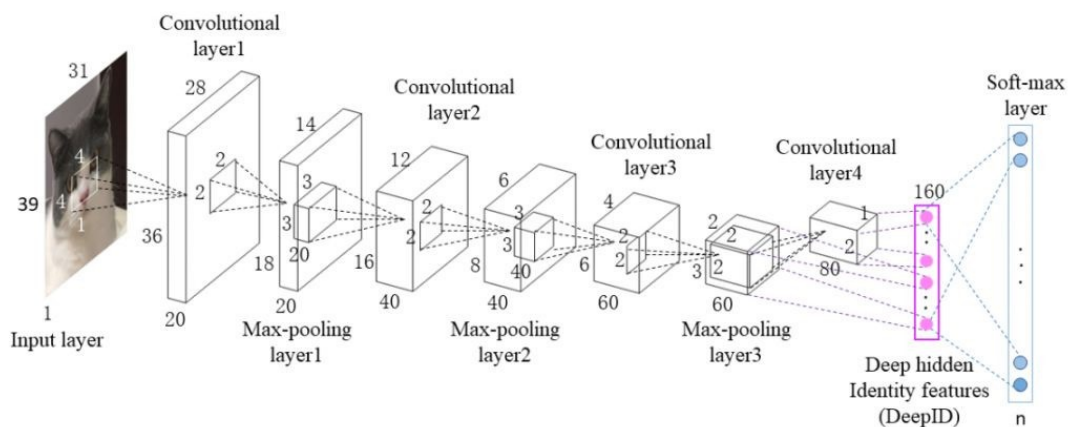


图 3.4 CNN 的结构示例

CNN 是深度学习中的重要模型，通过多层卷积和池化，能够自动提取不同层次的特征，参数共享和局部连接的设置减少了参数量，显著提高计算效率，CNN 广泛应用于图像处理和计算机视觉领域，在图像分类、目标检测和图像分割等任务中表现出色。

## 四、数据预处理与统计分析

### 4.1 附件 1 数据预处理

附件 1 中的数据为问卷数据为问卷数据，其中包括受访者的个人基本信息、ASD 相关量表数据、以及响应变量 PCL 总分。本文借助 python 的 numpy、pandas 库对数据进行处理，并结合 seaborn、matplotlib 库进行数据可视化分析。

#### 1、缺失值及异常值处理

首先，查看数据中是否存在缺失值。如果存在缺失值，需要根据具体情况对缺失值数据进行填充或者剔除缺失数据。借助 info() 命令得到数据缺失情况如图 5.12 所示。

```
RangeIndex: 903 entries, 0 to 902
Data columns (total 30 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   序号                                903 non-null    object
1   年龄                                903 non-null    int64
2   性别                                903 non-null    int64
3   籍贯                                903 non-null    object
4   民族                                903 non-null    object
5   身高                                903 non-null    int64
6   体重                                903 non-null    int64
7   文化程度                            903 non-null    int64
8   婚姻状况                            903 non-null    int64
9   家庭人均月收入                    903 non-null    int64
10  近期是否出现意外事件              903 non-null    int64
11  是否有家庭成员或亲近朋友意外      903 non-null    int64
12  意外事件中是否目睹有人重伤        903 non-null    int64
13  意外事件及救援过程中，您是否接触或见过尸体 903 non-null    int64
14  您是否曾因为灾难现场的景象        903 non-null    int64
15  既往是否有精神病史                903 non-null    int64
16  近一月内是否使用药物              903 non-null    int64
17  是否吸烟（包括以前）              903 non-null    int64
18  吸烟状态                          903 non-null    int64
19  是否饮酒                          903 non-null    int64
20  ASD分离                            903 non-null    int64
21  ASD再体验                          903 non-null    int64
22  ASD回避                            903 non-null    int64
23  ASD警觉                            903 non-null    int64
24  ASD总分                            903 non-null    int64
25  ASD性质                            903 non-null    int64
26  PCL总分                            903 non-null    int64
27  心理韧性                          903 non-null    int64
28  BMI                                903 non-null    float64
29  遗传史                            903 non-null    int64
dtypes: float64(1), int64(26), object(3)
memory usage: 211.8+ KB
```

图 4.1 附件 1 数据缺失情况图

可以看出，数据中并不存在缺失值，因此无需特地进行缺失值处理。附件 1 数据中共有 903 个样本，30 个变量。其中，像性别、文化程度等为分类变量，在附件 1 数据说明中给出了详细介绍；而诸如身高、体重、ASD 分离等变量为连续型变量，响应变量为 PCL 总分。

通过对婚姻状况变量进行频数分析，如表 4.1所示。可以看出，“婚姻状态”变量中存在 4 个取值为其他的样本，考虑为异常变量，将其对应样本剔除。

表 4.1 婚姻状况频数统计表

变量	取值	含义	频数
婚姻状况	1	未婚	753
	2	已婚	146
	3	其他	4
合计			903

## 2、响应变量处理

对 PCL 总分的数据分布情况进行可视化，如图 4.2所示。

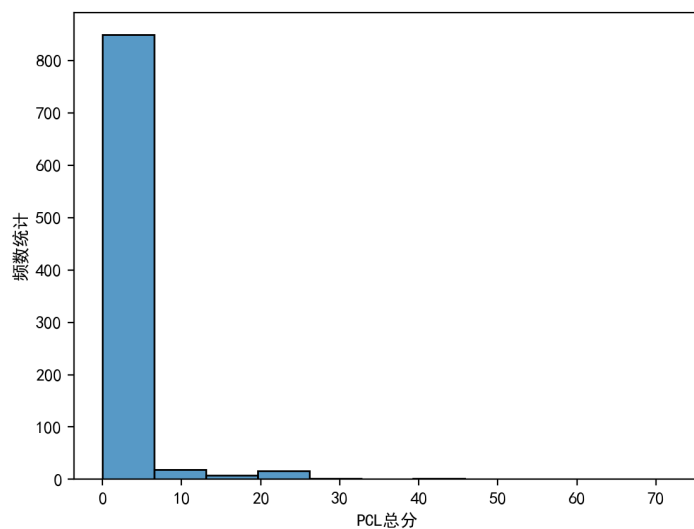


图 4.2 PCL 总分频数统计图

根据频数图分析可知，PCL 总分取值范围在 0-70 之间，但是却有超过 800 个样本的 PCL 总分集中在 0-5 之间，需要尝试对 PCL 总分数据进行转换。

通过查阅资料发现，PCL 总分是用来衡量 PTSD 发生风向的一个指标，其值是根据问卷数据中的相关问题得分计算得到的。根据问卷设置的不同，不同文献中对 PCL 总分的划分也有所不同。为此，本文按照 PCL 总分是否小于 3，对 PCL 总

分进行划分, 得到变量  $PCL_3$ 。当取值为 1 时, 我们认为其有较大可能发生 PTSD; 当取值为 0 时, 我们认为其不会发生 PTSD, 具体转换公式如下。

$$PCL_3 = \begin{cases} 0, & PCL\text{总分} \leq 2 \\ 1, & \text{else} \end{cases} \quad (4.1)$$

特征变换后的频数统计图如图 4.3所示。

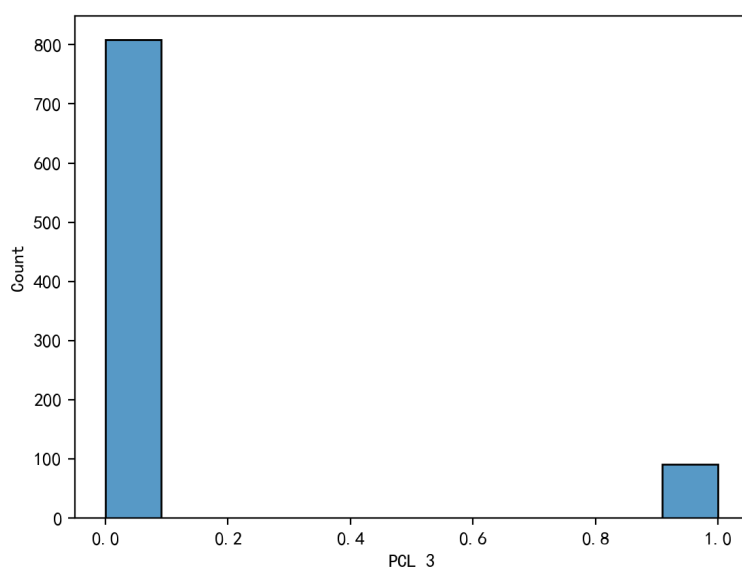


图 4.3 特征变换后的 PCL 频数统计图

此时, 我们将问题转换成了对  $PCL_3$  的二分类预测问题。同时, 结合表 4.2 具体分析, 数据存在轻度的不平衡性问题。

表 4.2 特征变换后的 PCL 频数统计表

	取值	频数	频率 (%)
$PCL_3$	0	812	89.92
	1	91	10.08

### 3、连续变量统计分析

针对数据中的连续变量, 对其进行简单描述统计分析, 然后, 绘制变量的分布柱状图, 直观查看变量的数据分布。连续变量的描述统计分析如表 4.3所示。

表 4.3 连续变量描述统计表

	年龄	身高	体重	ASD 分离	ASD 再体验	ASD 回避	ASD 警觉	ASD 总分	心理韧性	BMI
mean	24.21	174.19	71.10742	5.69	5.57	4.53	5.78	21.56	39.48	23.44
std	3.93	5.360	12.74	1.86	1.76	1.60	1.91	6.47	14.79	4.18
min	15	158	40	5	5	4	5	19	10	13
25%	22	170	64	5	5	4	5	19	36	21
50%	23	175	69	5	5	4	5	19	48	22.86
75%	26	178	75	5	5	4	6	21	50	24.21
max	56	200	130	21	25	20	25	91	50	45.35

接下来，绘制数据分布柱状图，如图 4.4、图 4.5所示。

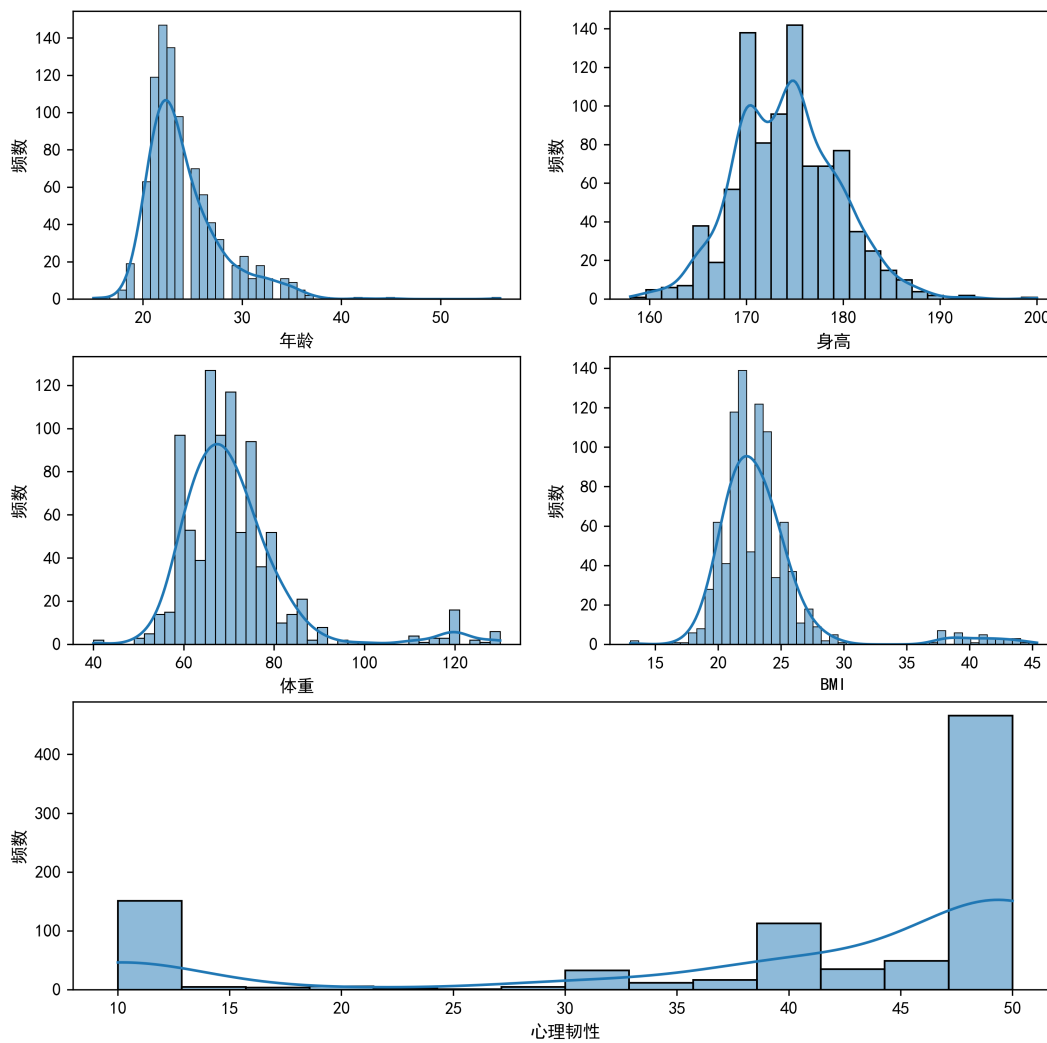


图 4.4 身高体重等指标柱状图

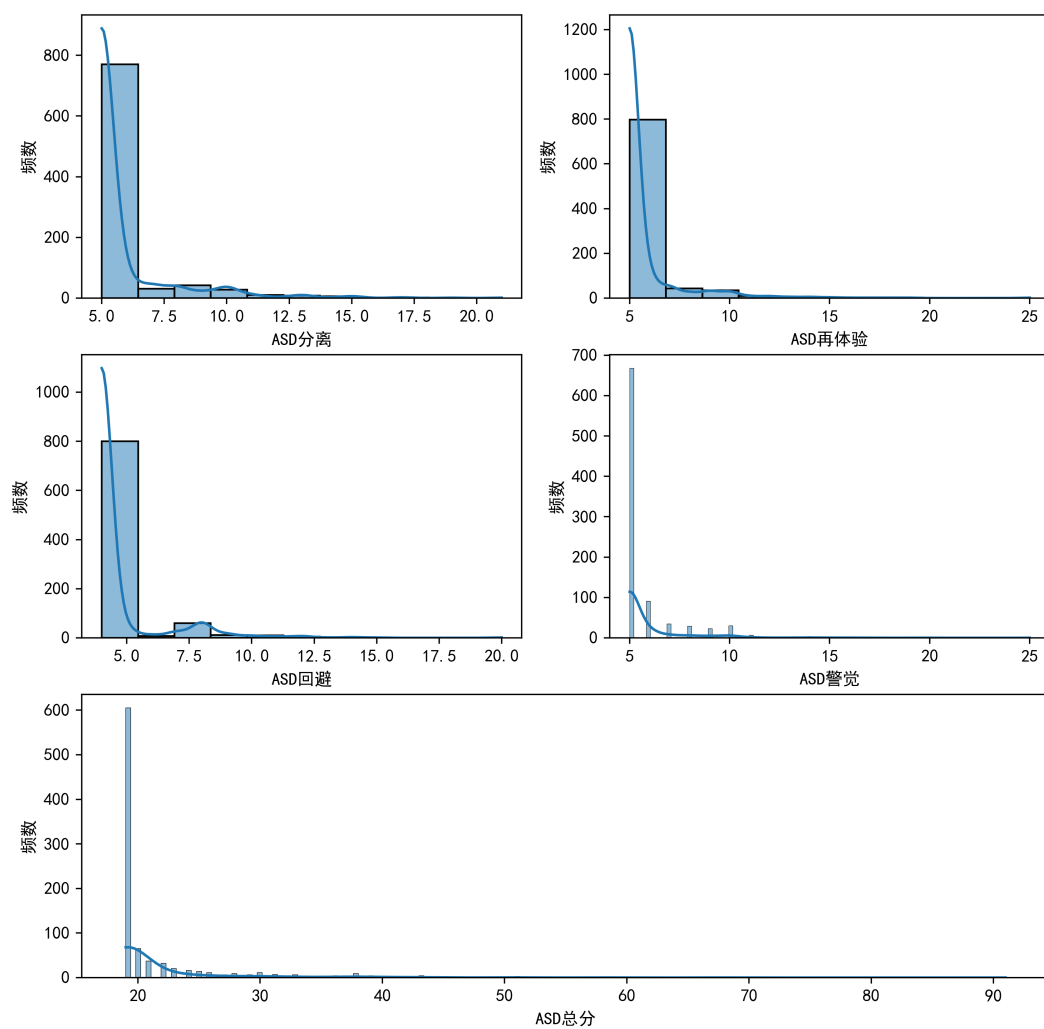


图 4.5 ASD 相关指标柱状图

#### 4、离散变量分析

针对数据中的离散变量，本文通过其统计频数表进行分析。如表 4.4所示。



表 4.4 离散变量频数统计表

变量	取值	含义	频数
性别	1	男	887
	2	女	12
文化程度	1	大学及以上	533
	2	高中	201
	3	中专及以下	165
婚姻状况	1	未婚	753
	2	已婚	146
家庭人均月收入	1	3000 以下	248
	2	3000-4999	321
	3	5000-8000	178
	4	8000 以上	152
近期是否出现意外事件	1	否	869
	2	是	30
是否有家庭成员或亲近朋友意外	1	否	876
	2	是	23
意外事件中是否目睹有人重伤	1	否	869
	2	是	30
意外事件及救援过程中，您是否接触或见过尸体	1	否	881
	2	是	18
您是否曾因为灾难现场的景象感到害怕、紧张或无助	1	否	825
	2	是	74
既往是否有精神病史	1	否	898
	2	是	1
近一月内是否使用药物	1	否	894
	2	是	5
是否吸烟（包括以前）	1	否	523
	2	是	376
吸烟状态	1	不吸烟，无被动吸烟	456
	2	不吸烟，被动吸烟	376
	3	吸烟	67
是否饮酒	1	否	855
	2	是	44
ASD 性质	0	阴性	867
	1	阳性	32
遗传史	1	阴性	896
	2	阳性	3

对表中数据进行分析发现,“是否有精神病史”变量只有 1 个取值为 1(是),“遗传史”变量只有 3 个取值为 2(阳性),“近一个月是否使用药物”变量只有 5 个取值为 2(是)。从信息熵的角度来看,这类变量所含有的信息较少,在实际建模中很难发挥作用,可以考虑将其进行初步剔除,后面结合其他方法进一步进行变量选择。

## 4.2 附件 1 数据特征提取

为了从数据中挖掘更多的信息,本文尝试对数据进行特征提取。将“籍贯”列按照“-”切分为“省份”和“城市”,共提取出 26 个省份、206 个城市,并按照省份所属地区进行特征提取。

在针对“民族”变量进行处理时,由于数据中多数为汉族,累计有 835 个样本,而其余民族的样本数累计仅有 64 个,故根据是否是少数民族进行特征提取。提取后的数据分布图如图 4.6 所示。

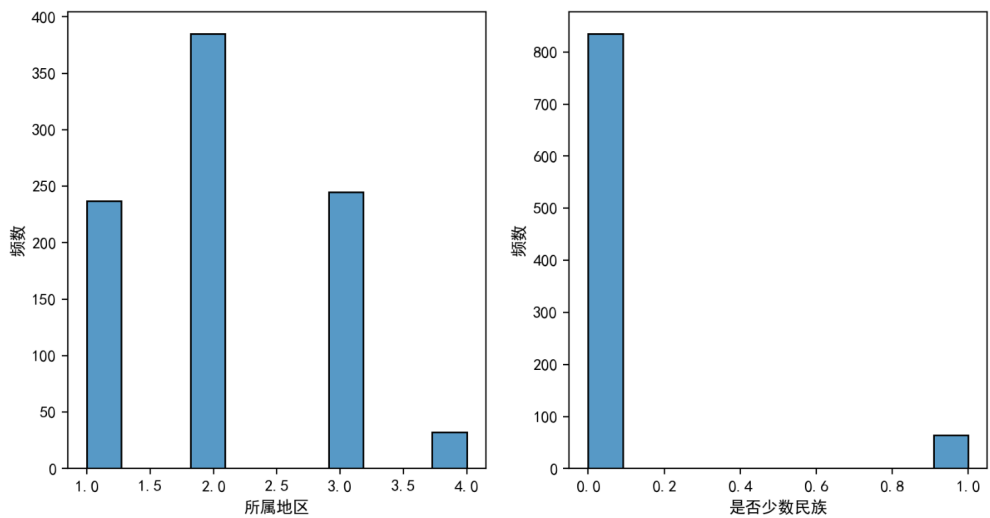


图 4.6 特征提取后的地区和民族柱状图

其中,所属地区柱状图中,1 代表“东部地区”,2 代表“中部地区”,3 代表“西部地区”,4 代表“东北地区”;是否少数民族图中,0 代表“汉族”,1 代表“少数民族”。

## 4.3 脑电波数据预处理

### 4.3.1 查看数据

附件 2 中 25 个脑电波数据文件，查看数据发现脑电连接信号异常统计如下图所示，其中文件 A9、A13、A16、A18、A19、A22、A23 的脑电连接信号异常个数均超过 20，其他信号连接异常较少。

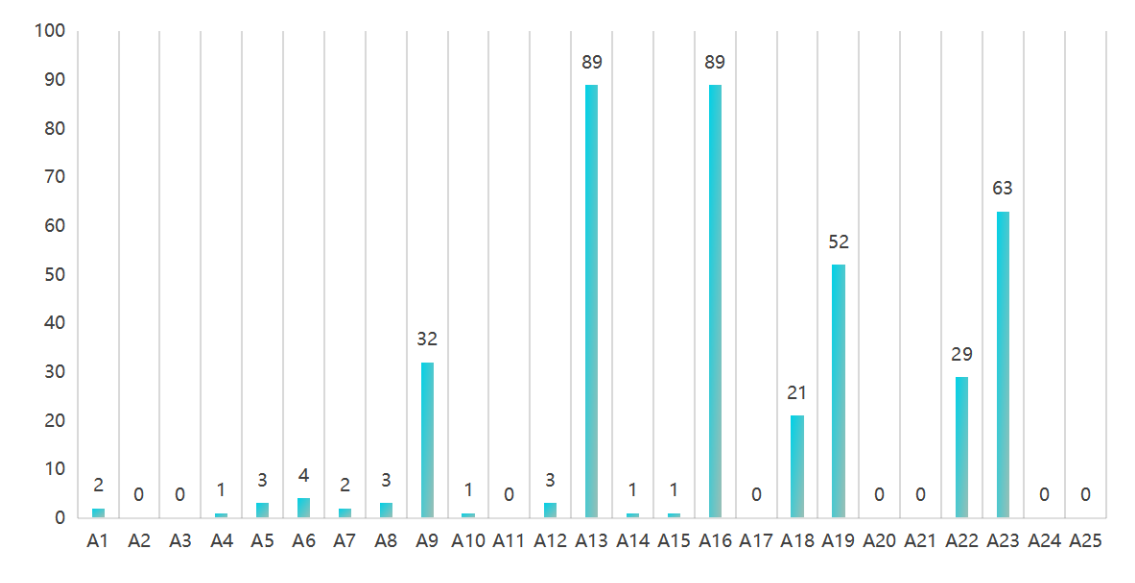


图 4.7 脑电波信号连接异常统计图

以样本 A1 中的十个变量为例，查看 A1-A25 样本中脑电波数据随时间的分布如下图所示。

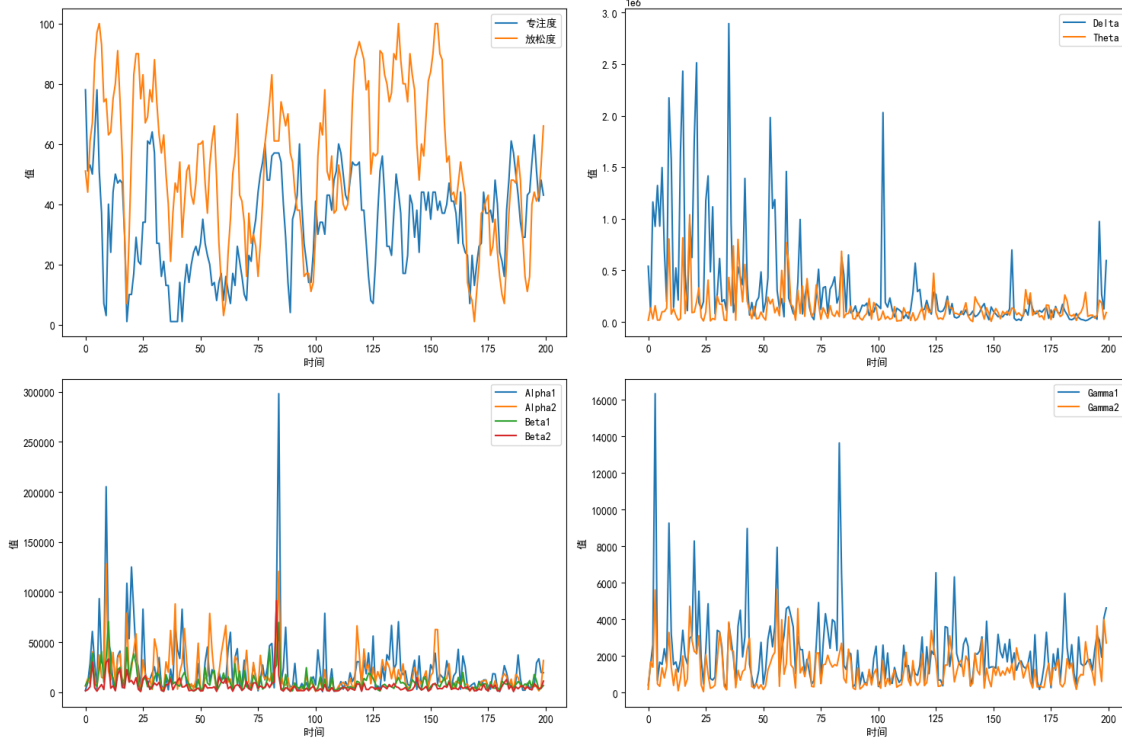


图 4.8 A1 预处理前脑电波数据分布图

从上图可以看出脑电波数据专注度、放松度、Delta、Theta、Alpha1、Alpha2、Beta1、Beta2、Gamma1 含有异常值。

#### 4.3.2 异常值处理

由于信号连接异常个数较多的 7 个文件异常值分布集中，本文不予考虑，仅考虑异常值较少的 18 个文件，统计发现信号连接异常的数据会突然变的巨大，不符合正常现象，因此本文采用三次样条插值，这是一种广泛应用于时间序列数据插值的方法。它使用三次多项式在每个区间上进行插值，并确保插值函数在节点处的一阶和二阶导数连续，使得插值曲线更加平滑。这种方法特别适用于平滑变化的时间序列数据。其基本思想是给定一组节点  $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ ，三次样条插值的目标是找到一组三次多项式  $S_i(x)$ ，使得每个多项式在相应的区间  $[x_i, x_{i+1}]$  上插值，并在节点处满足连续性条件。

每个区间上的三次多项式可以表示为：

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (4.2)$$

设  $h_i = x_{i+1} - x_i$ ，插值条件可表示为：

$$a_i = y_i \quad (4.3)$$

$$a_{i+1} = y_{i+1} \quad (4.4)$$

连续性条件和二阶导数连续性条件可表示为：

$$b_i + 2c_i h_i + 3d_i h_i^2 = b_{i+1} \quad (4.5)$$

$$2c_i + 6d_i h_i = 2c_{i+1} \quad (4.6)$$

边界条件（自然边界条件）可表示为：

$$c_0 = 0 \quad (4.7)$$

$$c_{n-1} = 0 \quad (4.8)$$

通过三次样条插值，可以在已知数据点之间构建出平滑且具有连续导数的插值曲线，插值后脑电数据如下图所示，可以看到已经过滤了个别异常值。

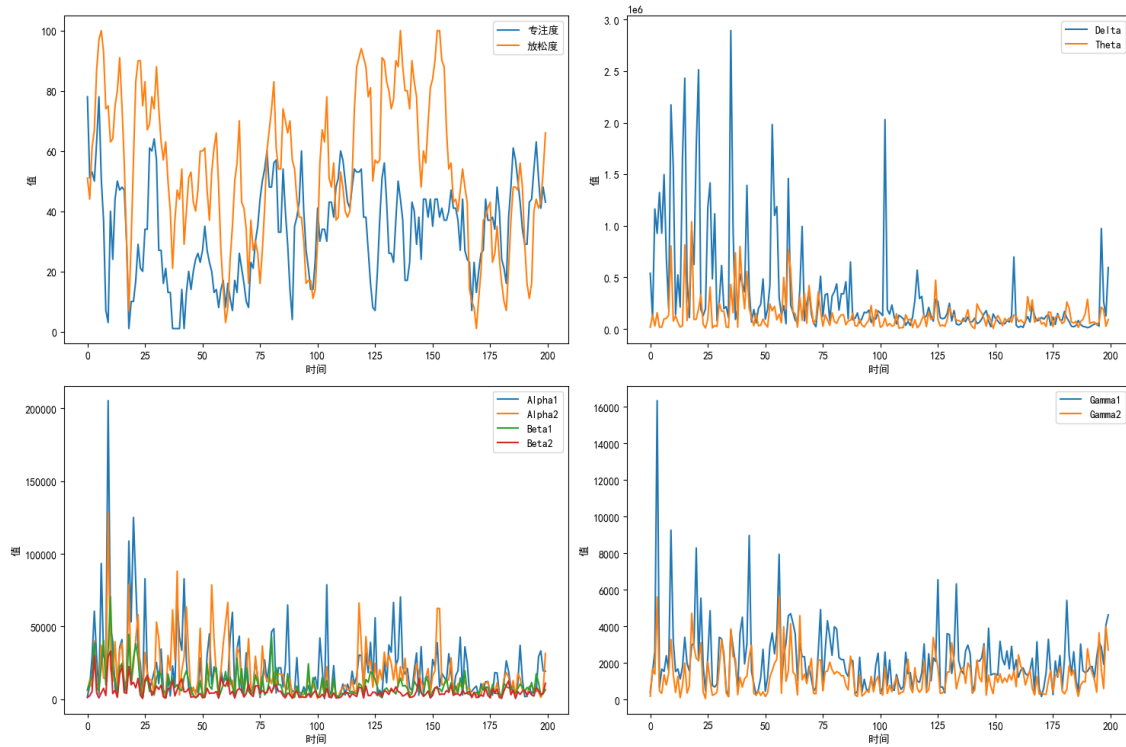


图 4.9 A1 插值处理后脑电数据分布图

### 4.3.3 数据去噪

由于脑电波 (EEG) 信号在采集过程中容易受到电力线干扰, 电极和皮肤接触不良或电极运动也会引起的噪声以及环境噪声等, 对脑电波数据进行滤波去噪是非常重要的, 本文通过 Butterworth 滤波器对脑电波数据进行去噪处理, 可以有效地去除低频和低频噪声, 仅保留感兴趣的频段, 从而提高信号质量和分析准确性。去噪后的数据分布如下图所示。



图 4.10 A1 滤波去噪后脑电数据分布图

对另外 18 个文件做同样的处理, 为以下建模做准备。

# 五、模型的建立与求解

## 5.1 基于 Logistic 回归的救援人员 PTSD 发病相关影响因素分析

本文结合相关文献，按照 PCL 总分是否小于 3 作为 PTSD 风险的判断标准，将对 PCL 总分的回归问题转化成了对 PCL 总分是否小于 3 的二分类问题。

针对数据中的多分类变量，进行具体分析。多分类变量有“文化程度”、“家庭人均月收入”、“吸烟状态”这三个，可以将这三个多分类变量看作等级变量，其中，“文化程度”是负向程度、“家庭人均月收入”和“吸烟状态”是正向程度。

### 5.1.1 Spearman 秩相关性分析

首先，本文针对  $PCL_3$  与自变量之间的相关性进行分析，由于  $PCL_3$  为类别变量，自变量中也有不少类类别变量，数据不服从正态分布，所以采用 Spearman 秩相关系数进行分析。得到相关系数热力图如图 5.1所示。

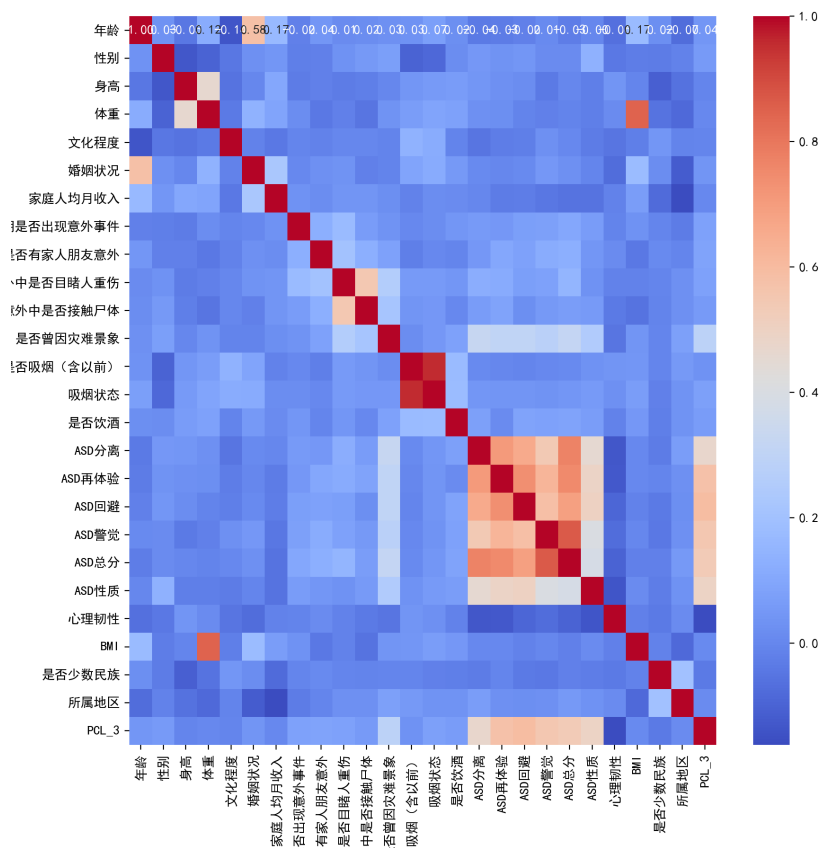


图 5.1 Spearman 相关系数热力图

从 Spearman 相关系数图中可以看出，与  $PCL_3$  相关性较高的变量有：ASD 分离、ASD 再体验、ASD 回避、ASD 警觉、ASD 总分、ASD 性质等；同时，这 6 个自变量之间又有较高的相关性，如果全纳入模型中，很有可能出现多重共线性。同时，BMI 为身高与体重的生成数据，在分析时也需要具体考虑。

### 5.1.2 Lasso-Logistic 回归建模

常见的用于解决二分类问题的模型有 Logistic 回归、决策树、支持向量机，以及集成方法随机森林和 Xgboost，考虑到要对 PTSD 发病的保护因素和危险因素进行分析，就要注重模型的可解释性。于是，本文选择 Logistic 回归进行建模求解。

同时，由于经过数据预处理、特征提取等一系列操作后，共保留了 27 个自变量，变量数目较多，可能会导致过拟合等问题。根据秩相关系数的分析结果发现，变量名中含有“ASD”的变量间具有较高的自相关系数，可能存在多重共线性等问题。而 Lasso 回归可以通过压缩回归系数的方式来进行变量选择，从而避免多重共线性、降低过拟合风险。

综上，我们使用带 L1 惩罚的 Logistic 回归进行建模分析，大致流程如图 5.2 所示。

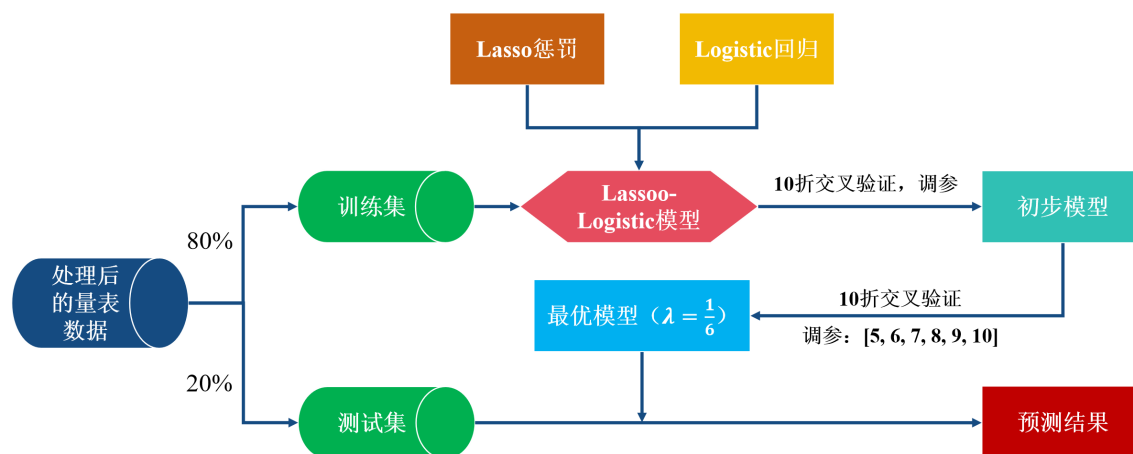


图 5.2 Lasso-Logistic 建模流程图

具体操作如下：

对保留的连续数据进行标准化处理，设置随机数 `random_state` 为 123，并按照 8：2 的比例切分训练集和测试集。我们使用训练集数据来拟合模型，为了减小



过拟合，我们在训练集上再采取 10 折交叉验证。同时，我们选择 L1 范数作为惩罚，来进行变量选择，确定  $\frac{1}{\lambda}$  的调参范围为  $[1, 2, 5, 10, 20, 50, 100]$ ，进一步缩小范围为  $[5, 6, 7, 8, 9, 10]$ ，来寻找最优的  $\frac{1}{\lambda}$ 。最终，确定的最优参数  $\frac{1}{\lambda}$  为 6，使用  $\frac{1}{\lambda} = 6$  重新建立模型，并在测试集上进行预测，以检验模型的预测效果和泛化能力，结果如图 5.3 所示。

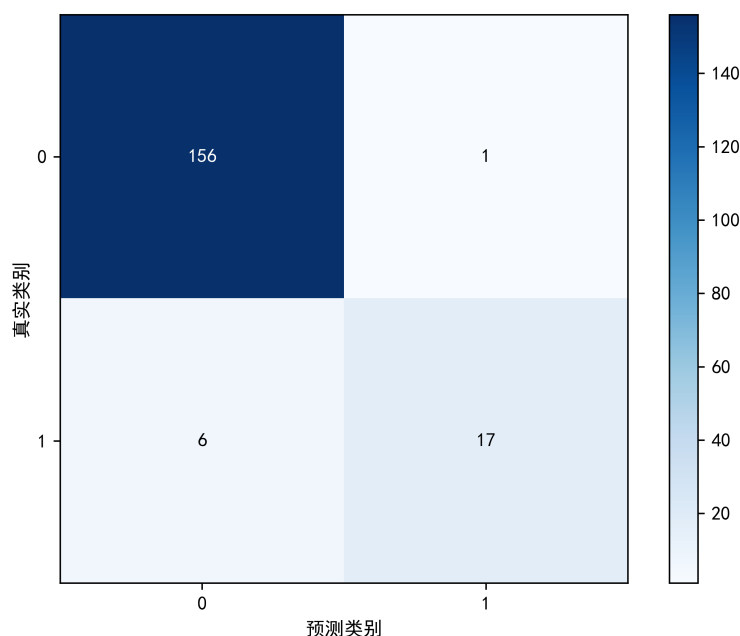


图 5.3 Logistic 回归在测试集上的混淆矩阵图

根据图 5.3，可以看出，测试集一共有 157 个标签为 0 的样本，模型预测对了 156 个，预测错了 1 个；一共有 23 个标签为 1 的样本，模型预测对了 17 个，预测错了 6 个，错误率较高。根据混淆矩阵图，计算模型评估指标 Accuracy、Precision、Recall、F1 值，具体如表 5.1 所示。

表 5.1 模型在测试集上的性能评估表

	Accuracy	Precision	Recall	F1 值
得分	0.96	0.94	0.74	0.83

根据表 5.6 可知，模型在测试集上的准确率为 0.96，精确率为 0.94，召回率为 0.74，F1 值为 0.83。模型在测试集上的准确率和精确率较高，而召回率较低，主

要是因为将不少原本属于类别 1 的样本分类到了类别 0 中。我们将在下一章中尝试进行优化。由于数据存在不平衡性，下面再结合 ROC 曲线及 AUC 值进行模型预测效果的评价。绘制 ROC 曲线如图 5.4所示。

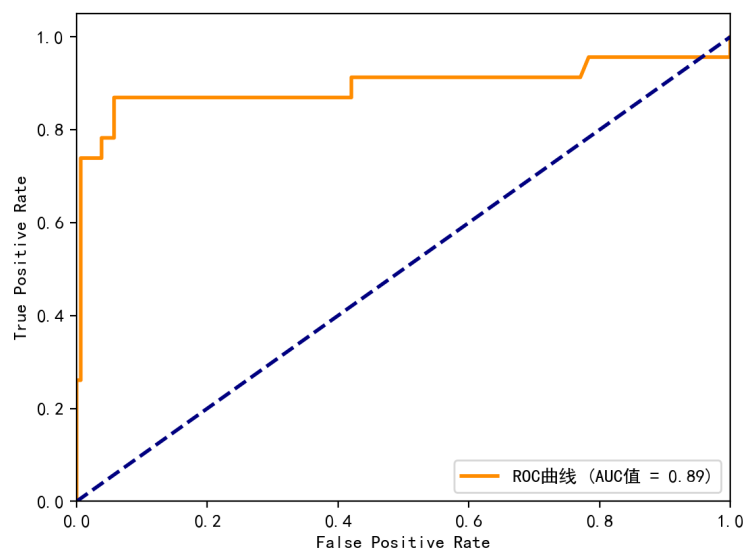


图 5.4 Logistic 回归在测试集上的 ROC 曲线图

从 ROC 曲线图中可以看出，ROC 曲线下面的面积 AUC 值为 0.89。结合表 5.1及图 5.4可以看出，Lasso-Logistic 模型对 PTSD 发病风险的预测效果较好。

### 5.1.3 救援人员 PTSD 发病相关影响因素分析

接下来，我们根据模型的回归系数对救援人员 PTSD 发病的保护因素和危险因素进行分析。由于本文采用的是 Lasso-Logistic 回归，可以进行变量选择和消除多重共线性。最终，选择参数  $\frac{1}{\lambda} = 6$  建立模型，保留的回归系数显著的变量及回归系数如表 5.2所示。

表 5.2 保留的变量及回归系数表

变量名	含义	回归系数
V1	年龄	0.0059
V2	文化程度	-0.0423
V3	家庭人均月收入	-0.2184
V4	意外事件及救援过程中, 您是否接触或见过尸体	-1.4724
V5	是否吸烟 (包括以前)	-0.1186
V6	是否饮酒	-0.4895
V7	ASD 再体验	0.0034
V8	ASD 回避	0.0254
V9	ASD 警觉	0.1440
V10	ASD 总分	0.0381

根据回归系数分析保护因素、危险因素及交互作用。由于因变量为 1 时代表发生 PTSD, 因变量为 0 时代表未发生 PTSD。所以, 当变量的回归系数为正数时, 随着变量值的增加, PTSD 风险也随之增加, 即为 PTSD 发病的危险因素; 当变量的回归系数为负数时, 随着变量值的增加, PTSD 风险反而减小, 即为 PTSD 发病风险的保护因素。

#### 1、对**保护因素**及作用机制进行具体分析如下。

(1) 家庭人均月收入, 即家庭人均月收入越高, PTSD 发病风险越低。救援人员也承担着养家糊口、照顾老人、抚养孩子的责任。家庭人均月收入越低, 可能会导致生活中的经济压力相对较大, 日常的精神状态相对较差。在面对创伤时, 相对而言会更多地考虑到家庭的生活重担, 担心自己如果不能工作家庭经济支出会难以为继, 会无形中增加创伤给自己带来的压力和刺激, 进而诱发 PTSD。

(2) 是否吸烟 (包括以前)、是否饮酒。即如果存在曾吸烟或饮酒的情况时, 可能会导致 PTSD 的患病风险降低。这与我们的先验知识和有关研究结论不符, 考虑可能与救援人员的职业特性或特殊的年龄结构有关。救援人员大多数以年轻人为主, 尤其附件 1 量表数据中, 受访者的年龄大多集中在 20-30 岁之间, 在时间上并无长期吸烟或饮酒的可能性存在。而吸烟或饮对 PTSD 的促进作用可能是较为长期才能体现出来。

#### 2、对**危险因素**及其风险机制进行分析如下。

(1) 年龄, 即随着年龄的增大, PTSD 的发病风险也随之增加。这与其他研究者的结论有些分歧。考虑原因可能如下。对救援人员的年龄绘制直方图可以显然,

样本中的救援人员年龄多数在 20 岁-30 岁之间，数据年龄分布并不均衡，有可能是在 20-30 岁这个年龄段之间，随着人生阅历的增加，逐渐成家立业，少有小，成为了家庭的顶梁柱，对危险的认识更加深刻，非常害怕自己出现危险，为家庭增加负担。在后续的深入分析时，可以考虑对年龄数据进行合适的特征提取，以进一步佐证分析。

(2) ASD 警觉，即随着 ASD 警觉值的增加，PTSD 发病风险也越高。ASD 警觉值越高，说明受访者出现焦虑、失眠、易怒、惶恐不安和高度警惕的程度越严重。显然，如果 ASD 警觉值越高，说明受访者对创伤后的应激反应越强烈，出现一系列心理和情绪问题，更容易导致诱发 PTSD。

(3) ASD 再体验、ASD 回避和 ASD 总分对 PTSD 发病而言也是危险因素。结合 ASD 的定义进行分析，ASD 一般是指受创后短期的应激反应，一般在受创后 4 周以内出现，如果不及时干预治疗，后期有转为 PTSD 的风险。

### 3、对交互作用进行分析如下。

通过交互作用分析，我们挖掘了以下存在显著影响因素的交互作用项，如表 5.3 所示。

表 5.3 显著的交互作用项

影响因素	交互项
保护因素	V1*V2; V1*V5; V1*V7; V1*V9; V1*V10; ...
危险因素	V2*V7; V4*V6; V5*V6; V3*V6; ...

(1) 针对是否吸烟（包括以前）、是否饮酒进行交互作用分析发现，在吸烟 + 饮酒的交互作用下，会促进发生 PTSD。这可能是由于仅吸烟或饮酒，可能会在短期缓解精神状况问题，但是同时吸烟 + 饮酒，可能会对精神状况有不良影响。

(2) 在意外事件中接触或见过尸体与是否饮酒之间的交互作用也会促使 PTSD 发生。这可能是由于受访者在意外事件中接触或见过尸体后，如果对其本身精神健康带来了影响，会通过饮酒来缓解。但是实际上，饮酒在遇到这种较强的刺激后，饮酒无益与缓解 PTSD 的发生。

(3) 为了探究不同收入群体的吸烟状态是否会造成 PTSD 发病风险的不同，进行交互作用分析。

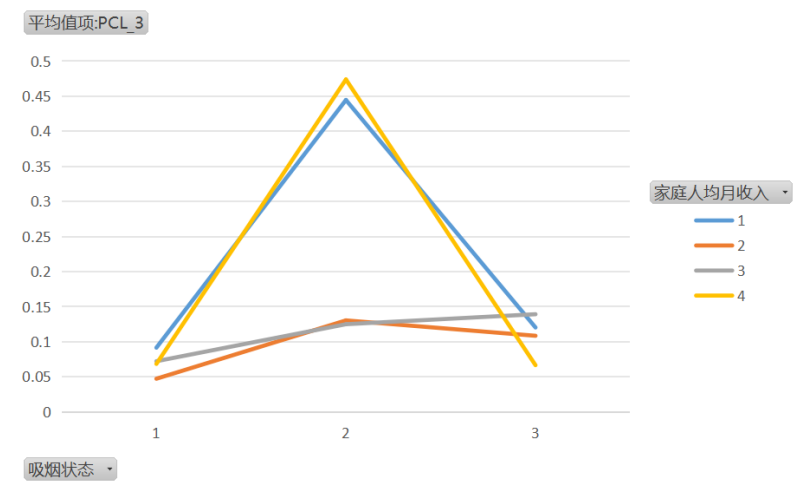


图 5.5 吸烟状态与家庭月收入交互作用图

从图 5.5中可以看出，不同收入群体的 PTSD 发病风险会随着吸烟状态的变化而变化。首先，针对无主动和被动吸烟、主动吸烟人群而言，不同收入不会导致 PTSD 发病风险的显著不同。而当吸烟状态为无主动吸烟但被动吸烟时，家庭人均月收入 3000 以下和 8000 以上的群体的 PTSD 发病风险更高，显著高于家庭人均男男女女月收入在 3000-8000 的群体。

(4) 为了探究不同婚姻状态的群体的 PTSD 发病风险是否随着学历的不同而变化，进行交互作用分析。

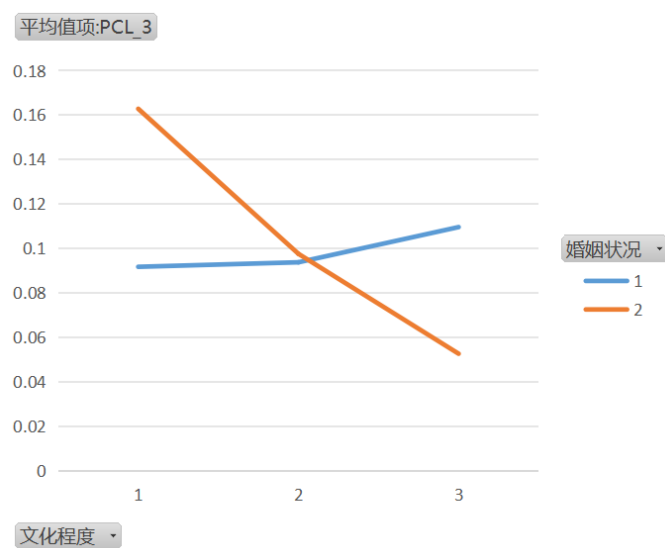


图 5.6 文化程度和婚姻状况交互作用图

---

从图 5.6中可以看出,高中学历人群,婚姻状况并没有导致 PTSD 发病风险有明显不同。随着文化程度的变化,不停婚姻状况的群体的 PTSD 发病风险也随之不同。横向来看,随着学历的增加,已婚群体的 PTSD 发病风险在快速增加,而未婚群体的 PTSD 发病风险则有减少趋势。纵向来看,针对本科及以上学历人群,结婚会促使 PTSD 发病风险增加;针对高中学历人群,区别不大;针对中专及以下人群,结婚则会促使 PTSD 发病风险降低。

## 5.2 基于多模态数据的 PTSD 风险评估模型

从第一问分析可知,附件 1 中家庭人均月收入、心理韧性、近期是否出现意外事件、是否有家庭成员或亲近朋友意外、您是否曾因为灾难现场的景象、吸烟状态、年龄、ASD 再体验、ASD 回避、ASD 警觉、意外事件中是否目睹有人重伤、意外事件及救援过程中是否接触或见过尸体共 12 个变量变量对是否引起救援人员 PTSD 有显著影响,这些变量包含连续数据和分类数据,而附件 2 中的脑电波是一系列时间序列数据,问题二旨在综合这些不同模态的数据来评估救援人员 PTSD 风险,比较和评估不同机器学习模型在脑电波数据和基本特征数据上的分类性能,以期找到最适合的分类评估模型。

### 5.2.1 数据融合原理

首先对连续数据标准化、分类数据编码、时间序列数据统计特征(如均值、标准差、最大值、最小值、偏度、峰度)。然后,通过早期融合、中期融合或晚期融合的方法,将这些特征组合成一个综合特征向量。最后,选择合适的机器学习或深度学习模型进行训练,并使用准确率、精确率、召回率等评估指标对模型进行评估。通过这种方法,可以全面利用不同模态的数据,具体结构如下。

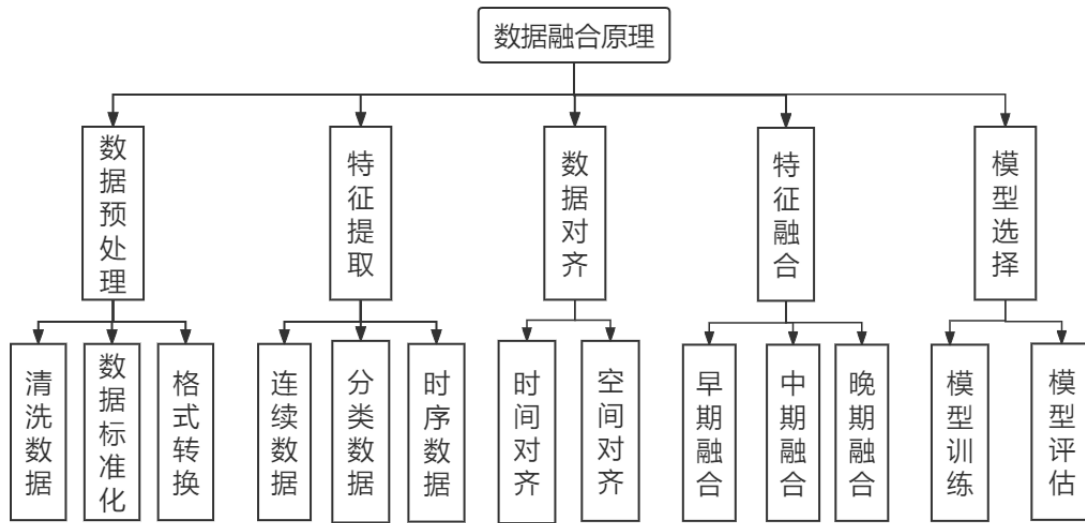


图 5.7 数据融合原理图

### 5.2.2 早期融合模型

早期融合，也称为特征级融合。在早期融合中，来自不同模态的数据在特征提取阶段就被融合在一起，然后作为一个整体输入到模型中进行训练和预测。步骤如下：

1. 读取附件 1 中显著变量特征数据和标签。
2. 对分类数据进行独热编码(One-Hot Encoding)和标签编码(Label Encoding)进行特征提取。
3. 提取脑电波数据统计特征包括均值、标准差、最大值、最小值、偏度和峰度。
4. 变量特征合并与标准化。

连续特征向量为  $\mathbf{x}_c$ ，分类特征向量为  $\mathbf{x}_d$ ，时间序列特征向量为  $\mathbf{x}_t$ ，则融合后的特征向量为：

$$\mathbf{x}_{\text{fused}} = [\mathbf{x}_c, \mathbf{x}_d, \mathbf{x}_t]. \quad (5.1)$$

标准化的公式如下：

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma}, \quad (5.2)$$

其中， $X$  是原始数据， $\mu$  是均值， $\sigma$  是标准差。

5. 数据集划分。
6. 模型训练与评估。

为了全面比较不同分类模型的性能，本文选择了三种常见的分类算法进行研究：支持向量机 (SVM)、多层感知器 (MLP) 和逻辑回归 (Logistic)。我们计算每个模型的准确率，并生成分类报告和混淆矩阵如下所示。

表 5.4 早期融合模型性能评估

	Accuracy	Precision	Recall	F1 值
SVM	0.43	0.55	0.43	0.45
MLP	0.71	0.86	0.71	0.73
Logistic	0.71	0.86	0.71	0.73

上表提供了准确率、精确率、召回率和 F1 分数等评估指标，可以看出早期融合模型 MLP 和 Logistic 回归的性能更好，三种模型的混淆矩阵则如下所示。

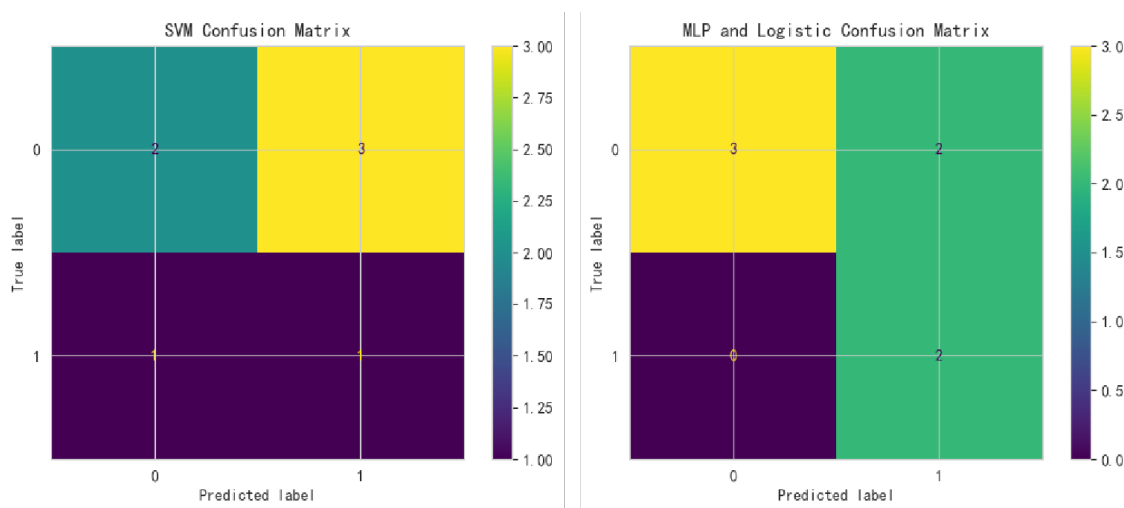


图 5.8 早期融合混淆矩阵

### 5.2.3 中期融合模型

中期数据融合是将来自不同数据源的特征在模型的中间层进行整合的方法。它先对每种数据源分别进行特征提取，然后在模型内部将这些特征进行拼接、加权平均或通过注意力机制融合，最后将融合后的特征输入到后续层进行进一步处理。

由早期融合知对于非时间序列特征 MLP 和 Logistic 回归的性能更好，为了充分利用脑电图的时间序列信息，在中期融合模式中，对于附件 1 的连续和分类数据我们采用 MLP 算法，脑电波数据分别采用 RNN、GRU 和 LSTM 三种时间序



列分析算法对每种数据源进行特征提取，将其转换为特征向量或特征张量。将融合后的特征输入到后续的神经网络（DNN）层进行处理。

步骤如下：

1. 数据加载与预处理：读取附件 1 显著特征数据和脑电波数据，截断时间序列数据以确保长度一致。将处理后的数据转换为 NumPy 数组并进行分割，得到训练集和测试集。

2. 模型构建：构建多层感知机（MLP）模型，用于处理附件 1 特征数据。构建 RNN（GRU 或 LSTM）模型，用于处理脑电波数据。构建（DNN）模型，用于融合多层感知机和 RNN（GRU 或 LSTM）的输出特征。

3. 融合模型：将 MLP 和 RNN 的输出特征拼接起来，通过 DNN 进行处理和输出最终的预测结果。

$$\mathbf{z} = \text{concat}(\mathbf{h}_{\text{MLP}}, \mathbf{h}_{\text{LSTM}}) \quad (5.3)$$

其中， $\mathbf{h}_{\text{MLP}}$  是 MLP 的输出， $\mathbf{h}_{\text{LSTM}}$  是 LSTM 的输出，concat 是特征拼接操作。

4. 训练和优化：将数据转换为 PyTorch 的 Tensor，使用二分类交叉熵损失函数（BCEWithLogitsLoss），计算预测值和真实标签之间的损失：

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\sigma(\hat{y}_i)) + (1 - y_i) \log(1 - \sigma(\hat{y}_i))] \quad (5.4)$$

其中， $\sigma$  是 sigmoid 激活函数， $y_i$  是真实标签， $\hat{y}_i$  是模型预测值。

反向传播和优化：使用 Adam 优化器，根据损失值计算梯度并更新模型参数：

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} \mathcal{L} \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} \mathcal{L})^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \\ \theta &= \theta - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \end{aligned} \quad (5.5)$$

其中， $m_t$  和  $v_t$  是动量和二阶动量估计， $\beta_1$  和  $\beta_2$  是动量衰减系数， $\alpha$  是学习率， $\epsilon$  是防止除零的小常数。

5. 特征提取和标准化：训练完成后，使用融合模型提取训练集和测试集的特征，对提取的特征进行标准化处理。

6. 模型评估：使用标准化后的特征训练逻辑回归模型，并进行预测。  
计算每个模型的准确率，并生成分类报告和混淆矩阵如下所示。

表 5.5 中期融合模型性能评估

	Accuracy	Precision	Recall	F1 值
MLP+RNN+DNN+Logistic	0.57	0.48	0.57	0.52
MLP+LSTM+DNN+Logistic	0.43	0.55	0.43	0.45
MLP+GRU+DNN+Logistic	0.86	0.88	0.86	0.84

上表可以看出中期融合模型 MLP 和 MLP+GRU+DNN+Logistic 组合的性能更好，该组合模型的混淆矩阵如下所示。

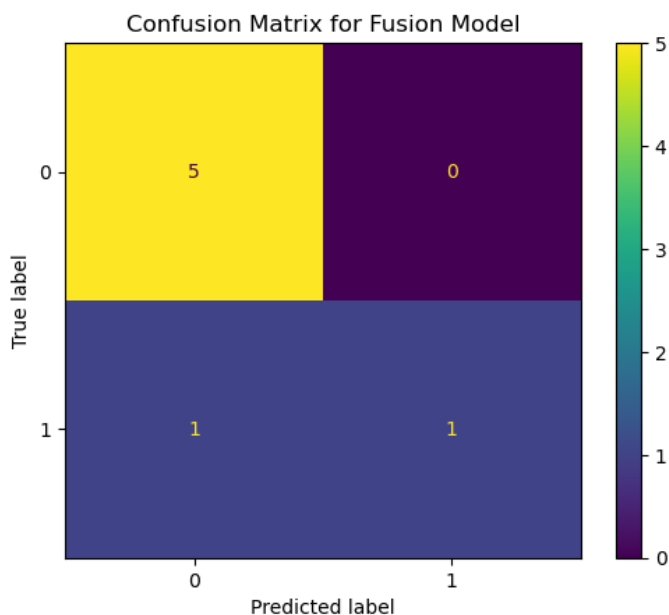


图 5.9 中期融合模型 GRU 混淆矩阵

#### 5.2.4 晚期融合模型

晚期数据融合（Late Data Fusion）是一种集成学习方法，用于结合来自多个模型的预测结果，以提高最终的预测性能。它是在各个模型独立完成其预测后，将这些预测结果进行融合或组合，形成最终的预测输出。卷积神经网络（CNN）可以自动提取有用的特征，捕捉局部模式，具有参数共享和计算高效性，能够处理平

移不变性，并且可以捕捉多尺度特征，从而提高模型的性能和鲁棒性，可以应用于时间序列分析。步骤如下：

- 1. 数据读取与预处理: 确保所有脑电波数据的形状一致，通过截断（取最短长度）处理。
- 2. 构建数据集和数据加载器: 自定义数据集类，拆分数数据集，创建数据加载器。
- 3. 定义模型 MLP 模型: 定义一个多层感知器（MLP）模型，用于处理附件 1 显著特征数据。GRU 模型: 定义一个门控循环单元（GRU）模型，用于处理脑电波数据。CNN 模型: 定义一个卷积神经网络（CNN）模型，用于处理脑电波数据。
- 4. 训练模型: 设置二进制交叉熵损失和 Adam 优化器。
- 5. 晚期融合。

模型评估: 在测试集上评估训练好的 MLP 和 GRU/CNN 模型，分别获取它们的预测结果。加权平均融合: 对两个模型的预测结果进行加权平均。

- 6. 结果可视化。
- 计算每个模型的准确率，并生成分类报告和混淆矩阵如下所示。

表 5.6 中期融合模型性能评估

	Accuracy	Precision	Recall	F1 值
MLP+GRU	0.71	0.71	0.71	0.71
MLP+CNN	0.43	0.55	0.43	0.45

上表 MLP+CNN 组合的效果并不好，可能是由于 CNN 模型虽然擅长捕捉数据中的局部模式和特征，但是更适合处理一维时间序列数据，而 GRU 模型更擅长捕捉时间序列数据中的时间依赖性和动态变化，因此性能更好。晚期融合模型的混淆矩阵如下所示。

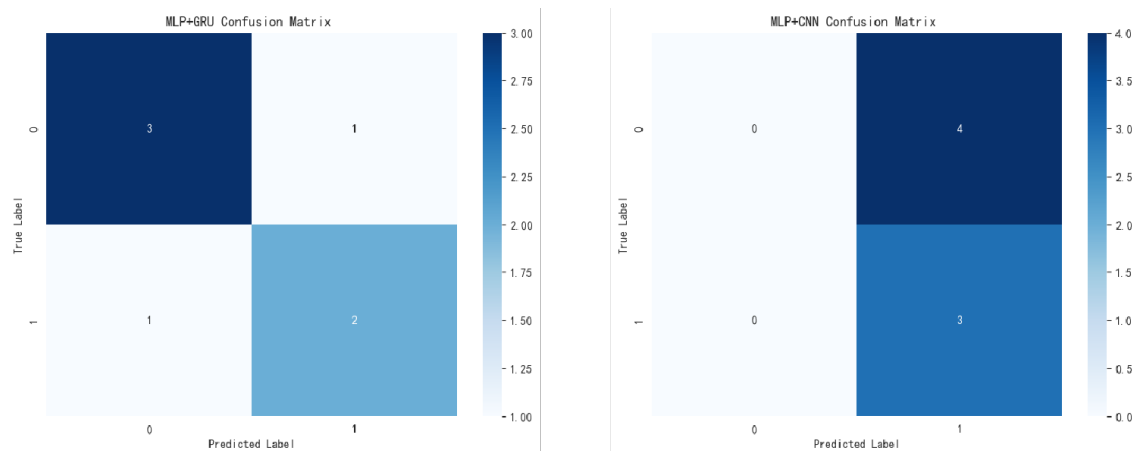


图 5.10 晚期融合模型混淆矩阵

### 5.2.5 个性化风险评估模型

在应急救援和灾难管理中，救援人员的生理和心理状态对任务的成功与否至关重要。提前预测和筛查救援人员的风险状态可以防止他们在不适合的情况下执行任务，从而提高救援的效率和安全性。

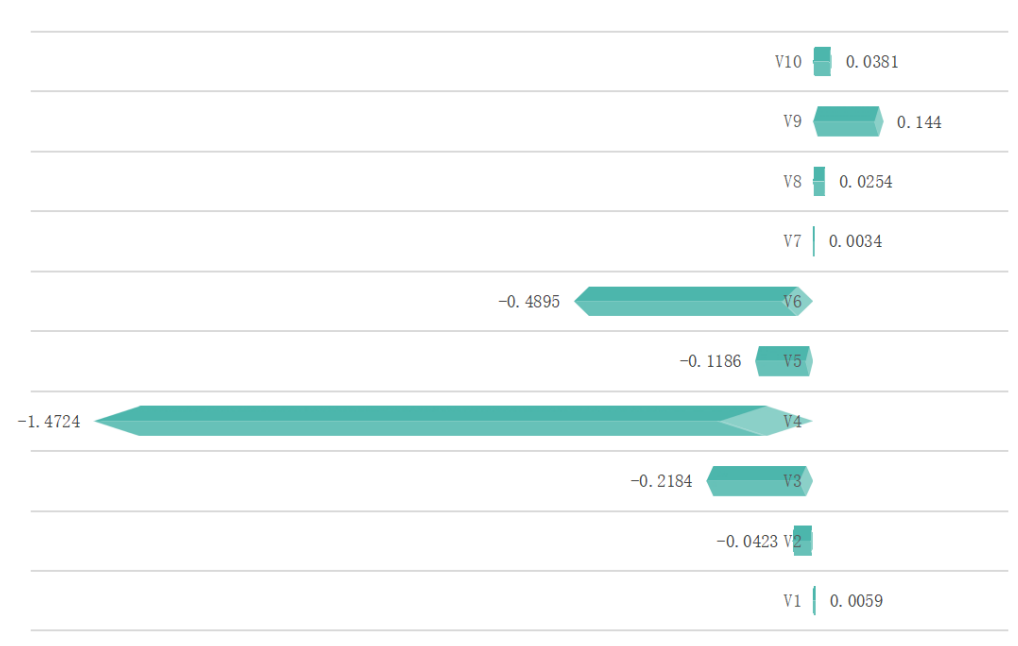


图 5.11 救援人员 PTSD 影响因素系数图

按照 PCL 总分是否小于 3 作为 PTSD 风险的判断标准，附件一中有 12 个显著的影响因素，其相关系数如上图所示，当变量的回归系数为正数时，随着变量

值的增加，PTSD 风险也随之增加，当变量的回归系数为负数时，随着变量值的增加，PTSD 风险减小，因此在对救援人员 PTSD 风险评估时应重点考虑收集以下 12 个因素。

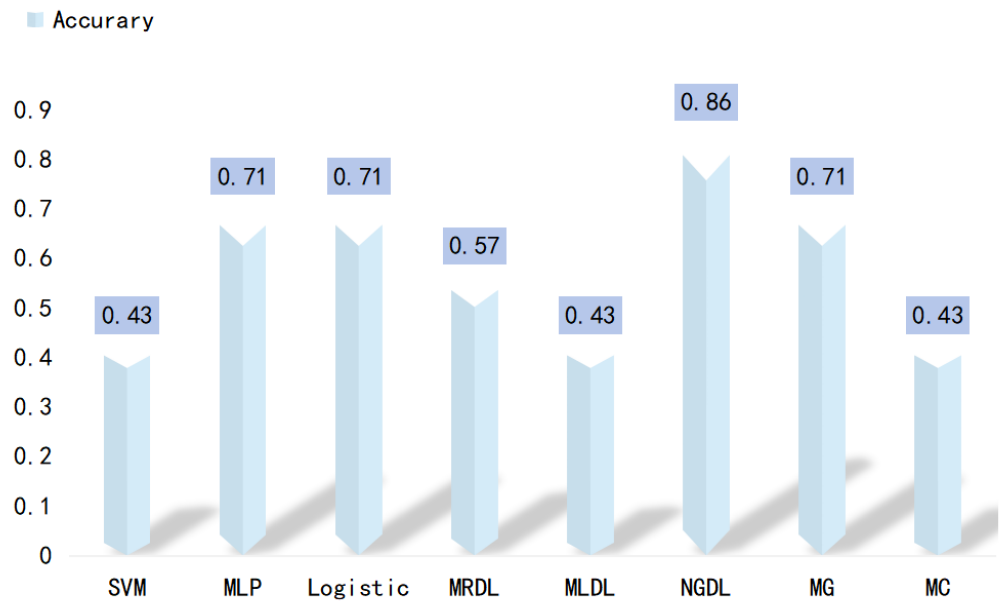


图 5.12 数据融合模型准确率

其次，根据以上分析，三种多模态数据融合模型的 Accuracy，如上图所示，中期数据融合组合模型 MLP+GRU+DNN+Logistic 的性能表现最优，即第一层采用多层感知机（MLP）模型用于处理附件 1 中显著影响的特征，GRU 模型用于处理脑电波的时间序列数据，将显著的连续变量、分类变量以及时间序列数据转化为相同维度的特征张量进行拼接，接着用拼接好的特征张量构建（DNN）模型，最后使用标准化后的特征构建逻辑回归模型进行预测评估，得到救援人员最终的 PTSD 患病概率，进而二分类判断或直接输出患病概率，以此实现利用多模态数据对救援人员是否患 PTSD 的风险评估，达到早期筛查，提前预防，降低风险的效果。

---

## 六、结论

### 6.1 结论

#### 1. 显著影响因素

救援人员的年龄、文化程度、家庭人均月收入、意外事件及救援过程中是否接触或见过尸体、是否吸烟（含以前）、是否饮酒、ASD 再体验、ASD 回避、ASD 警觉、ASD 总分共 10 个变量对是否引起救援人员 PTSD 有显著影响。

#### 2. 保护因素

家庭人均月收入，即家庭人均月收入越高，PTSD 发病风险越低。短期来看，吸烟或饮酒可能会降低 PTSD 的发病风险。

#### 3. 危险因素

年龄，随着年龄的增大，PTSD 的发病风险也随之增加。随着文化程度的增加，PTSD 的患病风险也在增加。危险因素还有：ASD 再体验、ASD 回避、ASD 警觉、ASD 总分，当 ASD 症状的相关指标值过高时，更要警惕其转化为 PTSD 的风险。

#### 4. 交互作用

家庭人均月收入越高，心理韧性得分相对较高，ASD 警觉相对较低，PTSD 发病风险相对较低。针对无主动和被动吸烟、主动吸烟人群而言，当吸烟状态为无主动吸烟但被动吸烟时，家庭人均月收入 3000 以下和 8000 以上的群体的 PTSD 发病风险更高，显著高于家庭人均男男女女月收入在 3000-8000 的群体。意外中目睹有人重伤、意外事件及救援过程中接触或见过尸体的救援人员对于灾难现场的景象会感到害怕、紧张或无助，PTSD 发病风险相对较高。

5. 多模态数据的 PTSD 风险评估模型设计针对分类数据、连续数据、脑电波时间序列数据的多模态数据融合模型设计，本文从早期融合 (SVM、MLP、logistic)，中期融合 (MLP+RNN+DNN+logistic、MLP+LSTM+DNN+logistic、MLP + GRU + DNN + logistic)，晚期融合 (MLP+GRU、MLP + CNN) 三个特征融合阶段出发，设计了 8 种方案并比较其性能，结果显示中期融合 MLP+GRU+DNN+logistic 组合对评估救援人员 PTSD 患病风险的效果最佳，由此设计了个性化的评估流程用于早期筛查预防。

---

## 6.2 建议

1、救援人员由于职业的特殊性，在进入一些危险场所展开救援时，本身自己也有受伤的风险。尤其是在展开救援过程中，如果有遇到有人重伤甚至接触尸体之后，要在救援结束后，重点关注这类人群的心理健康，必要的话开展专门的心理辅导或治疗，以避免发生创伤后应激障碍。同时，对救援人员个人生活经历中的一些意外事件也要保持关注。

2、适当提高救援人员的收入，于此同时，更为重要的是，要解决救援人员在开展救援工作过程中发生意外的善后和家庭保障问题。可以帮助救援人员缓解因家庭生活带来的压力，进而减少 PTSD 的发生。

3、针对 20-30 岁的救援人员，随着年龄的增加，越有可能在遇到创伤后发生 PTSD，为此，我们尤其要对 25-30 岁左右的年轻人，在其遭受创伤后进行重点关注。针对中专及以下学历且未婚的群体，大学及以上人群且结婚的群体，其发生 PTSD 的可能性较大，要保持重点关注（这里也有可能是反应出了从事救援工作的工作年限，随着工作年限的增加，在一定时间限制内，PTSD 的发病风险是增加的）。

4、在工作环境中，尽量给救援人员创造无“烟”的环境。如果有吸烟习惯，请逐渐戒除，同时在日常生活中远离二手烟。尤其是收入在 3000 以下和 8000 以上的群体，更要远离身边的二手烟。

5、针对救援人员，由于救援过程中会出现各种危险或精神刺激，需要救援人员有较好的心理韧性和较好的承压能力。可以定期组织救援人员开展跑步、游泳等集体健身运动，帮助其释放压力，同时锻炼抗压能力。

6、针对救援人员，定期开展心理健康调查与健康状况监测，如发现 ASD 症状，及早干预治疗，避免 PTSD 的发生。

---

## 参考文献

- [1] 李昌俊, 贾东立. 救援人员的主要心理问题、相关因素与干预策略[J]. 灾害学, 2021, 36(01): 148-152.
- [2] 伍芳慧, 刘媛. 创伤后应激障碍诊治研究进展[J]. 创伤外科杂志, 2021, 23(06): 472-475.
- [3] Grasser J, L.R. A. Treatments of Posttraumatic Stress Disorder in Civilian Populations[J]. Curr Psychiatry Rep, 2019, 21(11). DOI: 10.1007/s11920-019-0994-3.
- [4] Mitchell T F. Post-traumatic stress disorder in firefighters: Prevalence and associated factors[J]. Journal of Occupational Health Psychology, 2018, 23(1): 17-29.
- [5] Del Prado M D. PTSD prevalence among police officers: A systematic review [J]. Psychological Services, 2019, 15(3): 243-255.
- [6] McGuire J R. Military personnel and PTSD: A comprehensive review of prevalence and treatment[J]. Military Medicine, 2019, 184(4): e207-e214.
- [7] Zhang Y, Ji B, Li N, et al. Prevalence and predictors of PTSD, depression and anxiety in healthcare workers during the COVID-19 pandemic: A cross-sectional study in China[J]. BMJ Open, 2020, 10(10): e041967.
- [8] 李艳芳, 钟仲英, 李欢婷, 等. 严重创伤患者创伤后应激障碍现状及影响因素分析[J]. 临床普外科电子杂志, 2023, 11(01): 57-63.
- [9] 石岩, 杨罗东, 张子维, 等. 经历亲人亡故 PTSD 患者认知功能损害的研究[J]. 心理月刊, 2024, 19(04): 34-36+45. DOI: 10.19738/j.cnki.psy.2024.04.008.
- [10] 罗娟, 伍娟, 尹金芝, 等. 主动脉夹层术后患者发生创伤后应激障碍现状及影响因素研究[J]. 中国当代医药, 2023, 30(26): 9-12.
- [11] 周守志, 尹弘青, 金学勤, 等. 急诊科护士创伤后应激障碍与生存质量及睡眠质量的相关性研究[J]. 护士进修杂志, 2020, 35(15): 1357-1361. DOI: 10.16821/j.cnki.hsjx.2020.15.003.
- [12] 朱白鹭, 刘欢, 钟清玲, 等. 突发公共卫生事件中一线医护人员的心理弹性、创伤后应激障碍与生活质量相关性研究[J]. 沈阳医学院学报, 2024, 26(01): 43-47+52. DOI: 10.16753/j.cnki.1008-2344.2024.01.008.
- [13] 孔嘉文, 江俊颖, 刘志伟, 等. 消防救援人员创伤后应激障碍影响因素研究[J]. 消防科学与技术, 2021, 40(01): 122-125.
- [14] Wu Y, Dai Z, Jing S, et al. Prevalence and influencing factors of PTSD symptoms among healthcare workers: A multicenter cross-sectional study during



- 
- the surge period of the COVID-19 pandemic since December 2022 in the Chinese mainland[J/OL]. *Journal of Affective Disorders*, 2024, 348: 70-77. <https://www.sciencedirect.com/science/article/pii/S0165032723014763>. DOI: <https://doi.org/10.1016/j.jad.2023.12.008>.
- [15] 裘懿琳. 新冠肺炎疫情下中高风险地区居民创伤后应激障碍现状及影响因素分析[J]. *黑龙江人力资源和社会保障*, 2022(04): 48-51.
- [16] Chu X, Dai X, Yuan P, et al. Epidemical trends and risk factors of PTSD in parents of critically ill children: Evidence from both meta-analysis and subgroup analysis[J/OL]. *Journal of Affective Disorders*, 2024, 344: 242-251. <https://www.sciencedirect.com/science/article/pii/S0165032723012107>. DOI: <https://doi.org/10.1016/j.jad.2023.10.032>.
- [17] 黄欢, 杨佳佳, 孙利, 等. 脑卒中患者发生创伤后应激障碍的影响因素及其风险预测列线图模型构建[J]. *实用心脑血管病杂志*, 2021, 29(12): 35-39.
- [18] 张津, 刘川源, 海祥, 等. 外伤性骨折并发创伤后应激障碍的相关因素及其预测价值分析[J]. *大医生*, 2023, 8(03): 98-101.
- [19] 邓傲骞, 杨燕贻, 李云静, 等. 用机器学习算法预测长沙消防员患创伤后应激障碍的风险[J]. *中南大学学报 (医学版)*, 2023, 48(01): 84-91.
- [20] Zhang J, Sami S, Meiser-Stedman R. Acute stress and PTSD among trauma-exposed children and adolescents: Computational prediction and interpretation[J/OL]. *Journal of Anxiety Disorders*, 2022, 92: 102642. <https://www.sciencedirect.com/science/article/pii/S0887618522001153>. DOI: <https://doi.org/10.1016/j.janxdis.2022.102642>.
- [21] Horwitz A, McCarthy K, House S L, et al. Intensive longitudinal assessment following index trauma to predict development of PTSD using machine learning[J/OL]. *Journal of Anxiety Disorders*, 2024, 104: 102876. <https://www.sciencedirect.com/science/article/pii/S0887618524000525>. DOI: <https://doi.org/10.1016/j.janxdis.2024.102876>.