

基于机器学习的大学生心理问题早期发现研究

张 铭

天津仁爱学院信息与智能工程学院, 天津

收稿日期: 2025年2月17日; 录用日期: 2025年5月12日; 发布日期: 2025年5月27日

摘 要

本研究旨在通过机器学习算法实现大学生心理问题的早期发现。研究采用自编问卷和心理教师访谈的方式, 对天津某高校1000名学生的心理状况进行评估, 将其分为健康和亚健康两类。运用BP (Back Propagation)神经网络和极限学习机(ELM)算法分析问卷数据, 构建预测模型, 进而实现早期预警。结果显示, 基于BP神经网络的模型在测试集200名学生中, 准确识别出186名, 总准确率达92.7%, 其中亚健康识别率达到91.01%, 显著优于极限学习机(ELM)模型。研究表明, BP神经网络在识别大学生心理亚健康方面具有较高的准确性和可靠性, 为高校心理健康筛查提供了创新高效的新方法。

关键词

机器学习, 心理健康, 早期发现

Research on Early Detection of College Students' Psychological Problems Based on Machine Learning

Ming Zhang

School of Information and Intelligent Engineering, Tianjin Ren'ai College, Tianjin

Received: Feb. 17th, 2025; accepted: May 12th, 2025; published: May 27th, 2025

Abstract

The study aims to achieve early detection of psychological issues among university students through machine learning algorithms. Using self-designed questionnaires and interviews with psychological counselors, the psychological status of 1,000 students at a university in Tianjin was assessed and classified into two categories: healthy and sub-healthy. Back Propagation (BP) neural network and Extreme Learning Machine (ELM) algorithms were employed to analyze the questionnaire data and

construct predictive models, thereby enabling early warning. The results showed that the model based on the BP neural network accurately identified 186 out of 200 students in the test set, achieving an overall accuracy of 92.7%, with a sub-healthy identification rate of 91.01%. This significantly outperformed the Extreme Learning Machine (ELM) model. The study demonstrates that the BP neural network has high accuracy and reliability in identifying sub-healthy psychological states among university students, providing an innovative and efficient new method for mental health screening in universities.

Keywords

Machine Learning, Mental Health, Early Detection

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着社会的迅速发展和竞争压力的不断加剧,大学生群体面临着前所未有的心理挑战,心理健康已成为评定大学生整体素质的重要指标之一。心理亚健康状况,已经成为高校学生群体中普遍存在的现象。前人研究发现,大学生心理健康问题的出现不仅影响到学生的学习成绩和人际关系,更严重的可能导致心理障碍、心理疾病。如果不进行及时治疗,心理问题可能加重至中度或重度,极大影响学生的学习和生活,严重者可能出现自残或自杀倾向,严重影响其未来的社会适应和职业发展。因此,如何有效识别和预测大学生的心理健康状况,成为了教育学、心理学以及相关领域研究的一个重要课题(苏悦, 2021)。

机器学习是人工智能的重要分支,旨在让计算机自动从数据中学习知识,无需依赖明确的编程。通过算法解析数据模式进行预测或决策。自 20 世纪 50 年代起,机器学习经历了从符号学习到深度学习的演变。随着计算能力和大数据的发展,21 世纪初机器学习迎来爆发式增长。如今,机器学习广泛应用于图像识别、自然语言处理、医疗诊断、金融预测等领域,大力推动了各行业的智能化发展(贾绪计, 2021)。

心理问题的早期发现一般在症状出现后的 1 至 3 个月内进行,此时患者可能出现情绪低落、焦虑或睡眠障碍等轻微不适,但尚未严重影响生活,也未达到专业治疗的程度。心理问题的早期发现及干预能有效防止问题恶化,帮助患者更快恢复心理健康,并减少长期治疗的需要(李虹, 2003)。

大学生心理健康的早期干预不仅能够帮助学生及时调整心理状态,还能为高校的心理健康教育提供科学依据。传统的心理健康评估主要依赖专业心理教师的主观评分和面谈,这种方式在数据量较大时难以做到全面性和准确性。随着信息技术的飞速发展,机器学习算法逐渐成为解决这一问题的新途径。其中 BP (Back Propagation)神经网络和极限学习机(Extreme Learning Machine, ELM)算法在心理健康数据分析中表现出了卓越的性能,这些算法在分类和预测任务中能够提供高效且精确的结果。因此,本研究构建基于大学生心理健康调查问卷的预测模型,实现对大学生心理健康问卷进行心理健康和亚健康的分类,为高校提供一种更加高效、系统的心理健康筛查和预警机制,拓宽高校对学生心理健康预警的新思路。

2. 研究资料与方法

2.1. 研究对象

2024 年 10 月对天津某高校学生进行心理健康状况问卷调查。调查问卷采取线上形式,运用类型抽

样的方法, 将该校全体本科生按学院分类, 并按入学年份分为四个级别(不包括专升本学生), 每个年级随机抽取 250 至 270 人。本次调查共计发放电子问卷 1036 份, 回收问卷 1009 份, 其中有效问卷 1000 份, 有效回收率为 96.53%。其中男生 530 人, 女生 470 人, 年龄在 17 至 22 岁, 样本的性别比例较为均衡。从样本学生的年级出发, 大一年级有效问卷 267 份, 大二年级有效问卷 253 份, 大三年级有效问卷 256 份, 大四年级有效问卷 224 份。最终, 经过大学心理健康教育中心确认, 1000 名学生中有 166 人被评定为心理亚健康, 835 人为心理健康。基于以上调查结果开展相关的建模分析研究。

2.2. 调查方法

自编调查问卷。调查问卷包括年龄、是否挂科(是用 1 表示, 否用 2 表示)、性别(男用 1 表示, 女用 2 表示)、是否为独生子(是用 1 表示, 否用 2 表示)、家庭情况(单亲用 1 表示, 非单亲用 2 表示)、年级、恋爱状况(恋爱用 1 表示, 单身用 2 表示)、是否为低收入家庭(是用 1 表示, 否用 2 表示)、近期是否经历应激事件(是用 1 表示, 否用 2 表示)和锻炼身体(坚持每周锻炼用 1 表示, 其他用 2 表示)。

2.3. 建立模型

本研究采用 BP 神经网络和极限学习机(ELM)算法对问卷数据进行分析。BP 神经网络是一种多层前馈神经网络模型, 采用误差反向传播算法进行训练, 目前广泛应用于教育、医疗、工业等多个领域。通过 BP 神经网络算法对问卷数据进行监督学习, 利用多层前馈神经网络结构, 并通过反向传播算法迭代更新网络权重, 从而使网络的输出逐步逼近目标输出。计算公式如下:

$$net_j = \sum_{i=0}^n w_{ij} x_i + b_j$$

式中: w_{ij} 是输入层第 i 个神经元到隐藏层第 j 个神经元的连接权重, x_i 是输入层第 i 个神经元的输入值, b_j 是隐藏层第 j 个神经元的偏置。

隐藏层第 j 个神经元的输出为:

$$h_j = f(net_j)$$

式中: f 是激活函数。

极限学习机(ELM)是一种基于前馈神经网络(FNN)的机器学习方法, 广泛应用于监督学习和非监督学习问题。该算法可快速准确地识别问卷数据中的不同类别, 筛选出关键特征, 提升模型性能, 从而提供高效的预测模型。计算公式如下:

$$net_{ij} = \sum_{k=1}^n w_{kj} x_{ik} + b_j$$

式中: w_{jk} 是输入层第 k 个神经元到隐藏层第 j 个神经元的连接权重, b_j 是隐藏层第 j 个神经元的偏置。

总准确率(Overall Accuracy, OA)衡量的是算法模型在整体样本上的分类准确性, 总准确率(OA)能够直观地反映算法模型对全体样本的识别能力。总准确率(OA)的计算公式如下:

$$OA = \frac{TP + TN}{TP + FN + FP + TN}$$

式中: TP (True Positive)为真阳率, TN (True Negative)为真阴率, FP (False Positive)为假阳率, FN (False Negative)为被错误地识别为假阴率。

2.4. 数据处理

采用 Microsoft Excel 2010 整理数据, 使用 Matlab 2024、SPSS22.0 和 UnscramblerX10.4 进行数据分

析, 采用 Origin2020 作图。

3. 结果与分析

3.1. 因素相关性分析

将年龄、是否挂科、性别、是否为独生子女、家庭情况、年级、恋爱状况、是否为低收入家庭、近期是否经历应激事件和身体锻炼情况 10 种因素与心理亚健康状况进行相关性分析(图 1)。

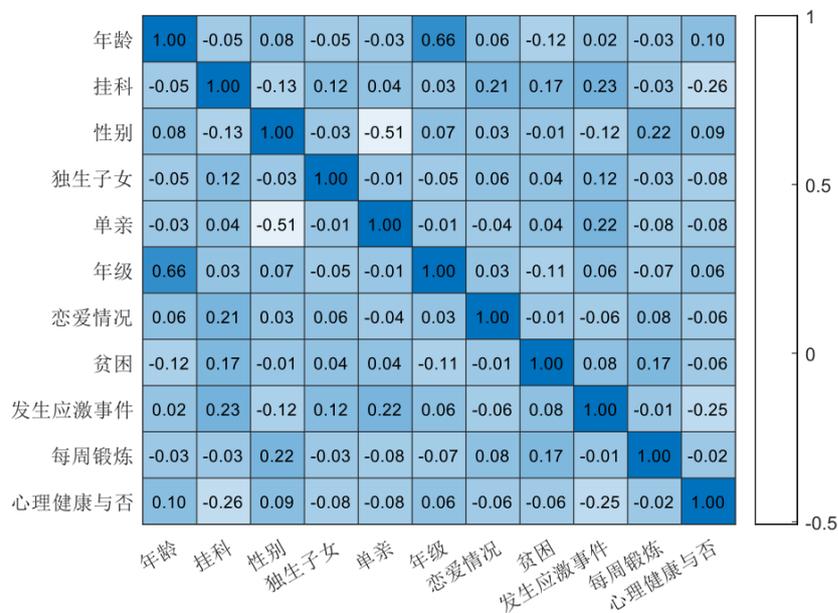


Figure 1. Correlation heatmap of various features
图 1. 各特征之间的相关性热图

其中, 年龄、性别、家庭结构和身体锻炼情况与心理亚健康之间的相关系数接近于 0, 这表明这些因素对心理亚健康的影响并不显著。这一结果证明这些因素并非心理健康的直接决定性因素, 而是通过其他中介变量间接影响心理健康。年龄可能通过影响生活压力和社会角色变化来间接影响心理健康; 性别可能通过社会期望和角色压力来发挥作用; 家庭结构可能通过家庭支持和关系质量来影响心理状态; 而身体锻炼情况可能通过影响身体状况和生活方式来间接作用于心理健康。

挂科情况与心理亚健康状况之间的相关系数为 -0.26 , 表明学业压力对心理健康有负面影响, 挂科可能导致学生产生焦虑、自卑等负面情绪, 进而影响心理健康; 家庭经济状况与心理亚健康状况之间的相关系数为 -0.25 , 表明经济状况对心理健康有负面影响, 低收入家庭的学生可能面临更多的经济压力和生活困境, 进而影响心理健康; 近期经历应激事件与心理亚健康状况之间的相关系数为 -0.25 , 表明应激事件对心理健康有负面影响, 经历应激事件可能导致学生产生焦虑、抑郁等负面情绪, 进而影响心理健康。

综上, 学业压力、家庭经济状况和近期应激事件是影响心理亚健康的重要因素, 而年龄、性别、家庭结构和身体锻炼对心理健康的影响无关。以上结果为制定心理健康干预措施提供了重要参考。

3.2. 准确率实证分析

本研究通过 BP 神经网络和极限学习机(ELM)两种算法模型, 对心理亚健康学生的识别效果进行了评估(表 1)。其中 1000 例学生样本用于训练模型, 200 例学生样本用于测试模型效果。基于 BP 神经网络模

型,亚健康识别率 91.01%,总准确率 92.7%,训练集准确率 93.65%,测试集准确率 91.64%。基于极限学习机(ELM)模型,亚健康识别率 78.84%,亚健康识别率 78.84%,训练集准确率 83.36%,测试集准确率 82.07%

BP 神经网络模型在真阳率和真阴率指标上均优于 ELM 模型。特别是在亚健康识别率和总准确率方面,BP 神经网络分别达到了 91.01%和 92.7%,而 ELM 模型仅为 78.84%和 82.65%。这表明 BP 神经网络在识别心理亚健康学生方面具有更高的准确性和可靠性(图 2)。

在测试集中,BP 神经网络模型正确识别了 185 例学生,准确率为 91.64%;其中,33 例心理亚健康学生中,正确判别了 30 例,准确率为 90.9%。综合来看,总准确率为 92.7%。结果表明,BP 神经网络算法在辅助筛查心理亚健康学生方面,具有高效性和精确性。

综上所述,本研究比较了两种经典机器学习模型 BP 神经网络和极限学习机(ELM)模型在心理健康预测上的能力,验证了 BP 神经网络在心理亚健康识别中的应用潜力,为心理健康筛查提供了新的方法和思路。未来研究可以进一步优化算法模型,提高识别准确率,并应用于更大规模的学生群体中,以期在实际教育环境中发挥更大的作用。

Table 1. Sample division table in training set and test set

表 1. 训练集与测试集样本划分表

算法模型	亚健康识别率(%)	总准确率(%)	训练集准确率(%)	测试集准确率(%)
BP 神经网络	91.01	92.7	93.65	91.64
ELM	78.84	82.65	83.36	82.07

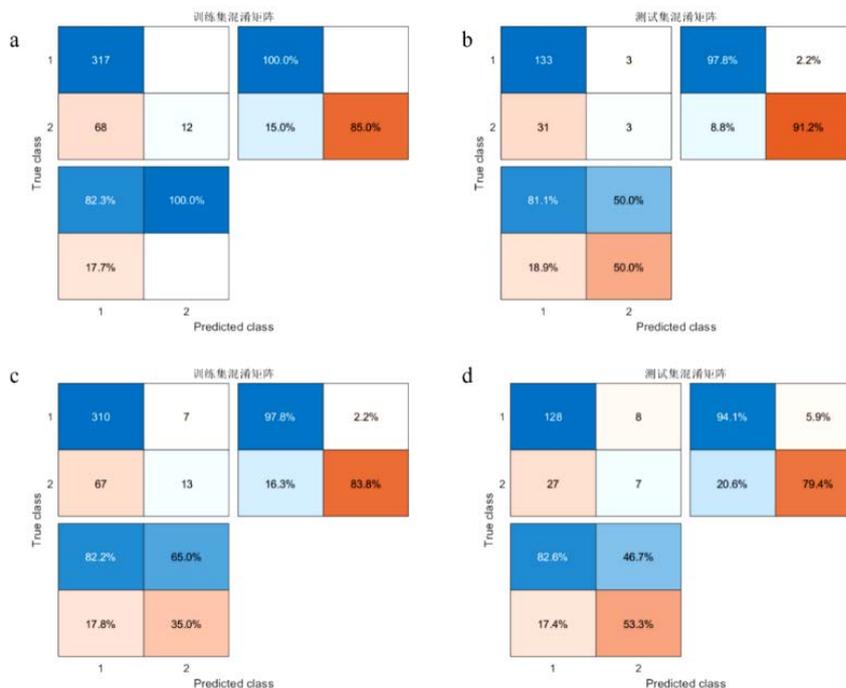


Figure 2. Correlation heat map of the confusion matrix for each feature: (a) Results of the training set for the BP neural network; (b) Results of the test set for the BP neural network; (c) Results of the training set for the ELM; (d) Results of the test set for the ELM

图 2. 各特征之间的相关性热图模型混淆矩阵:(a)为 BP 神经网络训练集结果;(b)为 BP 神经网络测试集结果;(c)为 ELM 训练集结果;(d)为 ELM 测试集结果

4. 讨论与结论

4.1. 讨论

本研究的重点在于利用机器学习算法对大学生心理健康状况进行早期识别和预警，为高校心理健康教育提供了科学依据，验证了机器学习算法在大学生心理健康筛查中的有效性。研究表明，BP神经网络模型在识别心理亚健康学生方面表现优异，其亚健康识别率和总准确率均高于极限学习机(ELM)模型，分别达到 91.01%和 92.7%。这表明 BP 神经网络在处理复杂数据关系和进行分类预测方面具有更高的准确性和可靠性。未来研究可以从以下方面进行深入探索。

4.1.1. 数据质量和维度

本研究使用的自编问卷虽然涵盖了多个可能影响心理健康的因素，但可能存在数据偏差或遗漏重要变量的情况。未来研究应进一步完善问卷设计，并考虑引入更多维度的数据，如生理指标、社交媒体行为等，以构建更全面的心理健康预测模型。还可以将模型应用于更大规模的学生群体，并探索其在不同人群、不同场景下的适用性，以提升模型的预测能力。

4.1.2. 模型可解释性

BP神经网络等深度学习模型虽然性能优异，但其“黑箱”特性使得模型决策过程难以解释。这可能导致用户对模型结果的信任度降低。未来研究应关注模型的可解释性，例如通过可视化技术或引入可解释性强的模型(如决策树)，以增强对模型的信任。还可以进一步优化机器学习算法模型，例如引入更多维度的数据、改进模型结构等，以提高预测准确率。

4.1.3. 伦理问题

机器学习在心理健康领域的应用涉及个人隐私和数据安全等问题。深入研究机器学习在心理健康领域应用的伦理问题，例如数据隐私、算法偏见等，如何在保护学生隐私的前提下，合理使用数据并确保模型结果的公平性，是未来研究需要重点关注的问题(刘任静, 2020)。

4.2. 结论

4.2.1. 基于 BP 神经网络的算法模型表现优异

在识别心理亚健康学生方面，BP神经网络模型的亚健康识别率和总准确率均高于极限学习机(ELM)模型，分别达到 91.01%和 92.7%。这表明 BP 神经网络在大学生心理亚健康识别方面具有更高的准确性和可靠性(徐洁, 2022)。

4.2.2. 模型具有应用价值

本研究验证了机器学习算法在大学生心理健康筛查中的应用潜力，为高校提供了一种高效、客观的心理健康预警方法。高校可以基于该模型建立动态监测和预警机制，实现对学生心理状态的实时跟踪和及时干预。

4.2.3. 影响因素分析

本研究发现，学业压力、家庭经济状况和近期应激事件是影响大学生心理亚健康的重要因素，而年龄、性别、家庭结构和身体锻炼对心理健康的影响相对较小。这一结果为高校制定心理健康干预措施提供了重要参考。高校应重点关注学业压力较大、经济困难以及近期经历应激事件的学生群体，并提供相应的心理支持和帮助。

综上所述，本研究基于机器学习算法，构建了大学生心理亚健康预测模型，并对其有效性进行了验

证。为大学生心理问题的早期发现和干预提供了新的思路和方法。未来,随着机器学习技术的不断发展,其在心理健康领域的应用前景将更加广阔。

基金项目

2024 年天津市高校思想政治教育工作研究基地支持(编号: JJSZY202407008); 2024 年天津市心理健康教育研究专项课题(编号: 2024GX25)。

参考文献

- 贾绪计(2021). 机器学习及其在心理健康领域的应用. *宁波大学学报(教育科学版)*, 43(4), 117-122.
- 李虹(2003). 大学生的压力与心理健康. *心理学报*, (2), 222-230.
- 刘任静(2020). 基于机器学习算法的大学生心理健康情况分析. *中国教育技术装备*, (22), 18-19+22.
- 苏悦(2021). 基于社交媒体数据的心理指标识别建模: 机器学习的方法. *心理科学进展*, 29(4), 571-585.
- 徐洁(2022). 基于 GA + BP 模型的大学生心理健康自动评级方法研究. *教育观察*, 11(32), 18-22+37.