

基于随机森林算法的中学生焦虑、抑郁影响因素分析及预测模型构建

高红霞^{1,2}, 陈靖^{1,2,3}, 张晓红^{1,2,3}, 侯丽明³, 王丽洁³, 孟亚清³, 高雯³, 黄亚欣³

1. 山西医科大学医学科学院, 山西 太原 030001; 2. 山西医科大学公共卫生学院流行病学教研室;

3. 山西省疾病预防控制中心学校卫生科

摘要:目的 基于随机森林算法分析 2023 年山西省中学生发生焦虑、抑郁情绪的影响因素并构建预测模型, 为防控中学生焦虑和抑郁等负面情绪的发生提供科学依据。方法 于 2023 年 9 月采用分层随机整群抽样方法, 在山西省 11 个市抽取 104 826 名中学生进行问卷调查, 使用广泛性焦虑量表(GAD-7)和流调中心抑郁量表(CES-D)调查焦虑抑郁情况。采用 Epidata 建立数据库, 组间比较采用卡方检验, 用 logistic 回归进行影响因素分析, 用 Python3.7 建立随机森林预测模型及自变量重要性排序。结果 2023 年山西省中学生焦虑和抑郁情绪检出率分别为 33.26% 和 23.72%, 女生高于男生, 普通高中和职高学生高于初中生。多因素回归分析结果显示, 女生(OR:1.846, 1.655, 95% CI:1.794~1.901, 1.602~1.710)、高中(OR:2.019, 1.670, 95% CI:1.956~2.083, 1.612~1.731)、职高(OR:1.167, 1.356, 95% CI:1.085~1.254, 1.254~1.466)、非常态家庭(OR:1.105, 1.248, 95% CI:1.072~1.140, 1.206~1.292)、住校(OR:1.102, 1.117, 95% CI:1.070~1.136, 1.080~1.155)、吃油炸食物 ≥ 1 次/天(OR:1.296, 1.375, 95% CI:1.227~1.369, 1.297~1.457)、受到严重伤害(OR:1.636, 1.688, 95% CI:1.524~1.755, 1.569~1.816)、遭受校园欺凌(OR:2.315, 3.411, 95% CI:2.150~2.492, 3.165~3.677)、被家长打骂(OR:2.881, 3.191, 95% CI:2.766~3.001, 3.059~3.328)、吸烟(OR:1.344, 1.435, 95% CI:1.266~1.427, 1.349~1.527)、饮酒(OR:2.062, 2.113, 95% CI:1.978~2.151, 2.022~2.208)、网瘾(OR:3.635, 3.976, 95% CI:3.289~4.016, 3.611~4.377)是中学生焦虑、抑郁情绪检出的危险因素;吃新鲜水果 ≥ 1 次/天(OR:0.824, 0.758, 95% CI:0.800~0.849, 0.733~0.783)、吃早餐频次高(有时吃:OR:0.843, 0.715, 95% CI:0.781~0.910, 0.661~0.773;天天吃:OR:0.538, 0.384, 95% CI:0.500~0.579, 0.356~0.414)、中高强度运动频次高(1~3天/周:OR:0.888, 0.821, 95% CI:0.854~0.923, 0.787~0.857; ≥ 4 天/周:OR:0.931, 0.848, 95% CI:0.893~0.971, 0.810~0.887)、体育课时数多(2~3节/周:OR:0.826, 0.859, 95% CI:0.801~0.853, 0.829~0.890; ≥ 4 节/周:OR:0.741, 0.758, 95% CI:0.695~0.789, 0.705~0.815)是焦虑、抑郁情绪检出的保护因素。喝含糖饮料 ≥ 1 次/天(OR:1.198, 95% CI:1.122~1.279)、吃蔬菜 ≥ 1 次/天(OR:0.861, 95% CI:0.828~0.895)分别是抑郁情绪检出的独立危险因素和独立保护因素。随机森林模型预测焦虑、抑郁情绪发生风险的 ROC 曲线下面积分别为 0.999 8、0.999 7, 灵敏度为 99.53%、99.67%, 特异度为 99.47%、98.98%。被家长打骂、网瘾和饮酒是中学生焦虑和抑郁情绪影响程度较大的共同因素。结论 山西省中学生焦虑、抑郁情绪检出率较高, 被家长打骂、饮酒、网瘾是焦虑和抑郁情绪发生影响较大的因素, 随机森林模型具有较高的预测价值。

关键词:焦虑; 抑郁; 影响因素; 随机森林; 预测模型

中图分类号: R749.72; R749.4 文献标志码: A 文章编号: 1003-8507(2025)02-267-08

DOI: 10.20043/j.cnki.MPM.202407151

Analysis of factors influencing anxiety and depression in secondary school students and construction of predictive model based on random forest algorithm

GAO Hong-xia*, CHEN Jing*, ZHANG Xiao-hong, HOU Li-ming, WANG Li-jie,

MENG Ya-qing, GAO Wen, HUANG Ya-xin

* Academy of Medical Sciences, Shanxi Medical University, Taiyuan, Shanxi 030001, China

Abstract: Objective To analyze the influencing factors on the occurrence of anxiety and depression among secondary school students in Shanxi Province in 2023 based on the random forest algorithm and to construct a prediction model to provide a

基金项目: 山西省卫生健康委员会卫生健康科研课题(2024155)

作者简介: 高红霞(1998—), 女, 硕士在读, 研究方向: 流行病学

通信作者: 陈靖, E-mail: 524210737@qq.com

scientific basis for preventing and controlling the occurrence of negative emotions such as anxiety and depression among secondary school students. **Methods** In September 2023, 104 826 secondary school students in 11 cities of Shanxi Province were selected by stratified random whole - cluster sampling method for questionnaire survey, and the Generalized Anxiety Scale (GAD - 7) and the Center for Streamlined Depression Scale (CES - D) were used to investigate anxiety and depression. Epidata was used for database establishment, chi - square test was used for between - group comparisons, logistic regression was used for influencing factor analysis, and Python 3.7 was used to establish a random forest prediction model and rank the importance of each variable. **Results** The detection rates of anxiety and depression among secondary school students in Shanxi Province in 2023 were 33.26% and 23.72%, respectively, and were higher for girls than for boys, and higher for students in general and vocational high schools than for junior high school students. The results of multifactorial regression analysis showed that female students ($OR: 1.846, 1.655, 95\% CI: 1.794 - 1.901, 1.602 - 1.710$), high school ($OR: 2.019, 1.670, 95\% CI: 1.956 - 2.083, 1.612 - 1.731$), vocational high school ($OR: 1.167, 1.356, 95\% CI: 1.085 - 1.254, 1.254 - 1.466$), non - conventional family ($OR: 1.105, 1.248, 95\% CI: 1.072 - 1.140, 1.206 - 1.292$), living in school ($OR: 1.102, 1.117, 95\% CI: 1.070 - 1.136, 1.080 - 1.155$), eating fried food ≥ 1 time/day ($OR: 1.296, 1.375, 95\% CI: 1.227 - 1.369, 1.297 - 1.457$), being seriously injured ($OR: 1.636, 1.688, 95\% CI: 1.524 - 1.755, 1.569 - 1.816$), being bullied in school ($OR: 2.315, 3.411, 95\% CI: 2.150 - 2.492, 3.165 - 3.677$), being scolded by parents ($OR: 2.881, 3.191, 95\% CI: 2.766 - 3.001, 3.059 - 3.328$), smoking ($OR: 1.344, 1.435, 95\% CI: 1.266 - 1.427, 1.349 - 1.527$), drinking alcohol ($OR: 2.062, 2.113, 95\% CI: 1.978 - 2.151, 2.022 - 2.208$), and Internet addiction ($OR: 3.635, 3.976, 95\% CI: 3.289 - 4.016, 3.611 - 4.377$) were the risk factors for the detection of anxiety and depressive mood in secondary school students, and eating fresh fruits ≥ 1 time/day ($OR: 0.824, 0.758, 95\% CI: 0.800 - 0.849, 0.73 - 0.783$), eating breakfast frequently (sometimes: $OR: 0.843, 0.715, 95\% CI: 0.781 - 0.910, 0.661 - 0.773$; daily: $OR: 0.538, 0.384, 95\% CI: 0.500 - 0.579, 0.356 - 0.414$), doing moderate to high - intensity exercise frequently (1 - 3 days/week: $OR: 0.888, 0.821, 95\% CI: 0.854 - 0.923, 0.787 - 0.857$; ≥ 4 days/week: $OR: 0.931, 0.848, 95\% CI: 0.893 - 0.971, 0.810 - 0.887$), and having a large number of physical education hours (2 - 3 sessions/week: $OR: 0.826, 0.859, 95\% CI: 0.801 - 0.853, 0.829 - 0.890$; ≥ 4 sessions/week: $OR: 0.741, 0.758, 95\% CI: 0.695 - 0.789, 0.705 - 0.815$) were the protective factors for the detection of anxiety and depressive protective factors for mood detection. Drinking sugary drinks ≥ 1 time/day ($OR: 1.198, 95\% CI: 1.122 - 1.279$) and eating vegetables ≥ 1 time/day ($OR: 0.861, 95\% CI: 0.828 - 0.895$) were independent risk factors and independent protective factors for the detection of depressed mood, respectively. The AUC of the Random Forest model predicting the risk of anxiety and depressed mood was 0.999 8 and 0.999 7, with a sensitivity of 99.53% and 99.67%, and a specificity of 99.47% and 98.98%, respectively. Being scolded by parents, Internet addiction and alcohol consumption were the common factors that had a greater degree of influence on anxiety and depression among secondary school students. **Conclusion** The detection rate of anxiety and depression among secondary school students in Shanxi Province is high, being scolded by parents, alcohol consumption and Internet addiction are the factors that have a greater influence on the occurrence of anxiety and depression, and the random forest model has a high predictive value.

Keywords: Anxiety; Depression; Influencing factors; Random forest; Prediction model

全球疾病负担研究显示,10~19岁学生中,约1/7存在精神障碍,占该年龄段总疾病负担的13%^[1],抑郁和焦虑分别是第四和第六位的致残原因^[2]。一项Meta分析表明我国中学生抑郁检出率为28.4%,高于欧洲、北美洲、亚洲发达国家^[3]。近三年研究表明我国中学生焦虑检出率为12.3%~34.6%^[4-6]。作为常见的精神疾病,焦虑、抑郁成为重大公共卫生问题,探讨其影响因素对早期预防和精准干预有重要意义。

有关中学生焦虑和抑郁情绪的影响因素已有研究报道,但对于其重要性排序及风险评估研究较少,且多采用传统的统计学方法探讨个体患病风险,易出现过拟合、预测效度低等问题。近年来,随机森林算法在疾病风险预测、预警和预后中广泛应用,而多因

素 logistic 回归能通过 OR 值对影响因素进行定量分析并作出直观解释,弥补随机森林算法解释困难的缺陷^[7]。本研究拟用随机森林和 logistic 回归相结合的方法分析 2023 年山西省中学生焦虑和抑郁情绪的现状及影响因素并构建预测模型,为识别高危人群和防控中学生焦虑和抑郁情绪的发生提供科学依据。

1 对象与方法

1.1 对象 于 2023 年 9 月采用分层随机整群抽样方法,抽取山西省 11 个市 117 个县(市、区)初中、普通高中和职业高中的一至三年级学生,按城区和郊县分层,每个城区 5 所中学(每区初中 2 所,高中 2 所,职高 1 所),每个郊县 3 所中学(每县初中 2 所,高中 1 所),每个年级抽取 2 个班级,确保每年级至少 80

名学生,以整班为单位开展调查。所有研究对象均签署知情同意书,研究经山西省疾病预防控制中心伦理审查委员会审核批准(审批号: SXCDCIRBYJ2024 093001)。本次调查共回收问卷 104 826 份,剔除信息不完整的问卷,得到有效问卷 103 877 份,有效回收率 99.1%。

1.2 方法

1.2.1 调查方法 采用国家编制的《2023 年全国学生常见病和健康影响因素监测与干预工作方案》中的“学生健康状况及影响因素调查表(中学版)”^[8]进行调查,问卷内容主要包括基本信息(学校类型、学段、学生年级、性别、家庭类型等)、饮食运动行为(过去一周喝含糖饮料、吃油炸食物、吃蔬菜水果、吃早餐频次、每天中高强度运动频次及体育课频次)、伤害相关行为(受到严重伤害、遭受校园欺凌及被家长打骂)、吸烟饮酒行为、网络使用情况(是否有网瘾)等。由经过统一培训的疾控中心人员担任调查员,通过学生自填问卷开展调查。

1.2.2 焦虑情绪 采用 Spitzer 等^[9]编制的广泛性焦虑量表(Generalized Anxiety Disorder, GAD-7)测量,该量表由 7 个条目组成,用来评估过去两周焦虑情绪的发生频率,每个条目按 0~3 分进行 4 级评估,选项 0~3 分别表示没有、有几天、一周有一半以上时间、几乎每天都有。总分 0~21 分,分数越高表明焦虑情绪越严重,总分 ≥ 5 分提示存在焦虑症状,分值 5、10、15 分别为轻、中、重度焦虑程度分界值。

1.2.3 抑郁情绪 采用 Radloff^[10]编制的流调中心抑郁量表(Center for Epidemiologic Studies Depression Scale, CES-D)测量,包括抑郁情绪、积极情绪、躯体症状与活动迟滞、人际共 4 个因素,按照过去 1 周出现相应情况的频度进行自评,采用 0~3 分 4 级计分制,没有或偶尔 = 0 分,有时 = 1 分,时常或一半时间 = 2 分,多数时间或持续 = 3 分,第 4、8、12、16 题为反向计分,总分范围为 0~60 分,得分 ≥ 16 分为出现抑郁风险的标准,得分越高抑郁程度越严重。

1.2.4 随机森林算法 随机森林是基于决策树的机器学习算法,它通过集成多个决策树对数据进行综合分类、关联性检验,将训练数据按照一定比例抽样,训

练一个决策树,通过投票或平均等方式得出最终的预测结果^[7]。

1.3 质量控制 由经过统一培训的调查员现场发放问卷,学生采用无记名方式独立填写问卷,完成后由调查员回收问卷并核查信息完整性。以双人双录入方式进行一致性检验。

1.4 统计学方法 采用 EpiData 建立数据库录入数据,用 SPSS 26.0 软件进行统计分析。计数资料采用构成比或率的形式进行统计描述,用 Pearson χ^2 检验比较焦虑、抑郁情绪检出率在不同人口学特征的组间差异,用非条件 logistic 回归进行影响因素分析,检验水准为双侧 $\alpha = 0.05$ 。按照 SMOTE 算法^[11]处理数据不平衡问题,用 Python 3.7 软件对自变量进行重要性排序,建立预测中学生发生焦虑和抑郁情绪的随机森林模型,选用灵敏度、特异度及受试者工作特征曲线(Receiver Operating Characteristic, ROC 曲线)下面积(Area Under Curve, AUC)评价预测模型的性能。

2 结果

2.1 一般人口学特征 本研究共纳入 103 877 名中学生,其中男生 52 239 人(50.29%),女生 51 638 人(49.71%);城区 22 085 人(21.26%),郊区 81 792 人(78.74%);初中生 62 823 人(60.48%),普通高中 36 185 人(34.83%),职业高中生 4 869 人(4.69%);来自常态家庭学生 75 319 人(72.51%),非常态家庭学生 28 558 人(27.49%);住校学生 57 412 人(55.27%),非住校学生 46 465 人(44.73%)。

2.2 中学生焦虑、抑郁情绪检出情况及单因素分析

2023 年山西省中学生发生焦虑情绪 34 546 人,焦虑情绪检出率为 33.26%,24 635 人有抑郁风险,抑郁情绪检出率为 23.72%。女生、高中、城区、非常态家庭、住校、喝含糖饮料 ≥ 1 次/天、吃油炸食物 ≥ 1 次/天、吃新鲜水果 < 1 次/天、吃蔬菜 < 1 次/天、从来不吃早餐、从来不做中高强度运动、一周上体育课节数、受到严重伤害、遭受校园欺凌、被家长打骂、吸烟、饮酒、有网瘾的中学生焦虑情绪和抑郁情绪检出率明显较高,差别均有统计学意义($P < 0.05$)。见表 1。

表 1 中学生基本情况以及焦虑、抑郁情绪检出情况[$n(\%)$]

Table 1 Basic information and detection of anxiety and depression in secondary school students[$n(\%)$]

变量	调查人数	焦虑情绪			抑郁情绪		
		检出人数 (检出率)	χ^2	P	检出人数 (检出率)	χ^2	P
性别			969.483	<0.001		343.729	<0.001
男	52 239(50.29)	15 009(28.73)			11 118(21.28)		
女	51 638(49.71)	19 537(37.83)			13 517(26.18)		
学段			2 904.002	<0.001		1 430.670	<0.001

(续表)

变量	调查人数	焦虑情绪			抑郁情绪		
		检出人数 (检出率)	χ^2	<i>P</i>	检出人数 (检出率)	χ^2	<i>P</i>
初中	62 823(60.48)	17 003(27.06)			12 364(19.68)		
高中	36 185(34.83)	15 851(43.81)			10 829(29.93)		
职高	4 869(4.69)	1 692(34.75)			1 442(29.62)		
学校位置			10.288	0.001		4.533	0.033
城区	22 085(21.26)	7 544(34.16)			5 357(24.26)		
郊区	81 792(78.74)	27 002(33.01)			19 278(23.57)		
家庭类型			99.887	<0.001		327.901	<0.001
常态家庭	75 319(72.51)	24 371(32.36)			16 754(22.24)		
非常态家庭	28 558(27.49)	10 175(35.63)			7 881(27.60)		
住校			528.509	<0.001		498.880	<0.001
否	46 465(44.73)	13 717(29.52)			9 497(20.44)		
是	57 412(55.27)	20 829(36.28)			15 138(26.37)		
喝含糖饮料			188.860	<0.001		522.195	<0.001
<1次/天	97 648(94.00)	31 979(32.75)			22 414(22.95)		
≥1次/天	6 229(6.00)	2 567(41.21)			2 221(35.66)		
吃油炸食物			565.024	<0.001		796.507	<0.001
<1次/天	96 075(92.49)	31 000(32.27)			21 765(22.65)		
≥1次/天	7 802(7.51)	3 546(45.45)			2 870(36.79)		
吃新鲜水果			687.117	<0.001		1 116.754	<0.001
<1次/天	40 945(39.42)	15 562(38.01)			11 949(29.18)		
≥1次/天	62 932(60.58)	18 984(30.17)			12 686(20.16)		
吃蔬菜			61.618	<0.001		409.488	<0.001
<1次/天	22 028(21.21)	7 813(35.47)			6 358(28.86)		
≥1次/天	81 849(78.79)	26 733(32.66)			18 277(22.33)		
吃早餐			1 634.485	<0.001		2 997.774	<0.001
从来不吃	3 629(3.49)	1 713(47.20)			1 576(43.43)		
有时吃	25 616(24.66)	10 740(41.93)			8 641(33.73)		
天天吃	74 632(71.85)	22 093(29.60)			14 418(19.32)		
中高强度运动			614.095	<0.001		784.884	<0.001
0天	17 755(17.09)	7 310(41.17)			5 643(31.78)		
1~3天	49 021(47.19)	15 716(32.06)			11 046(22.53)		
≥4天	37 101(35.72)	11 520(31.05)			7 946(21.42)		
体育课时数			876.420	<0.001		658.143	<0.001
0~1节	28 213(27.16)	11 290(40.02)			8 194(29.04)		
2~3节	68 517(65.96)	21 396(31.23)			15 131(22.08)		
≥4节	7 147(6.88)	1 860(26.02)			1 310(18.33)		
受到严重伤害			765.334	<0.001		953.936	<0.001
没有	99 869(96.14)	32 404(32.45)			22 869(22.90)		
有	4 008(3.86)	2 142(53.44)			1 766(44.06)		
遭受校园欺凌			1 096.171	<0.001		2 253.879	<0.001
没有	100 237(96.50)	32 411(32.33)			22 575(22.52)		
有	3 640(3.50)	2 135(58.65)			2 060(56.59)		
被家长打骂			3 796.898	<0.001		4 895.486	<0.001
没有	90 474(87.10)	26 952(29.79)			18 241(20.16)		
有	13 403(12.90)	7 594(56.66)			6 394(47.71)		
吸烟			1 682.796	<0.001		2 431.476	<0.001
否	96 979(93.36)	30 701(31.66)			21 316(21.98)		
是	6 898(6.64)	3 845(55.74)			3 319(48.12)		
饮酒			4 026.470	<0.001		4 575.562	<0.001
否	88 478(85.18)	26 001(29.39)			17 688(19.99)		
是	15 399(14.82)	8 545(55.49)			6 947(45.11)		
网瘾			1 989.317	<0.001		2 869.869	<0.001
否	101 415(97.63)	32 697(32.24)			22 934(22.61)		
是	2 462(2.37)	1 849(75.10)			1 701(69.09)		

注:()内数字为检出率/%。

2.3 中学生焦虑、抑郁情绪的多因素分析 依据 GAD-7 ≥ 5 分划分为焦虑组和非焦虑组, CES-D ≥ 16 分划分为抑郁组和非抑郁组, 将单因素分析中有意义的影响因素纳入多因素 logistic 分析。结果显示, 女生、高中、职高、非常态家庭、住校、吃油炸食物 ≥ 1 次/天、受到严重伤害、遭受校园欺凌、被家长打

骂、吸烟、饮酒、网瘾是焦虑、抑郁情绪检出的危险因素 (OR: 1.102 ~ 3.635, 1.117 ~ 3.976); 吃新鲜水果 ≥ 1 次/天、吃早餐频次高、中高强度运动频次高、体育课时数多是焦虑和抑郁检出的保护因素 (OR: 0.538 ~ 0.931, 0.384 ~ 0.859)。见表 2。

表 2 中学生焦虑、抑郁情绪的多因素 logistic 分析

Table 2 Logistic regression analysis of anxiety and depression among secondary school students

变量	焦虑情绪		抑郁情绪	
	OR(95% CI)	P	OR(95% CI)	P
性别(以男生为参照)				
女生	1.846(1.794 ~ 1.901)	<0.001	1.655(1.602 ~ 1.710)	<0.001
学段(以初中为参照)				
高中	2.019(1.956 ~ 2.083)	<0.001	1.670(1.612 ~ 1.731)	<0.001
职高	1.167(1.085 ~ 1.254)	<0.001	1.356(1.254 ~ 1.466)	<0.001
学校位置(以城区为参照)				
郊区	1.012(0.976 ~ 1.050)	0.505	1.040(0.999 ~ 1.084)	0.058
家庭类型(以常态家庭为参照)				
非常态家庭	1.105(1.072 ~ 1.140)	<0.001	1.248(1.206 ~ 1.292)	<0.001
住校(以否为参照)				
是	1.102(1.070 ~ 1.136)	<0.001	1.117(1.080 ~ 1.155)	<0.001
喝含糖饮料(以 <1 次/天为参照)				
≥ 1 次/天	1.010(0.949 ~ 1.074)	0.760	1.198(1.122 ~ 1.279)	<0.001
吃油炸食物(以 <1 次/天为参照)				
≥ 1 次/天	1.296(1.227 ~ 1.369)	<0.001	1.375(1.297 ~ 1.457)	<0.001
吃新鲜水果(以 <1 次/天为参照)				
≥ 1 次/天	0.824(0.800 ~ 0.849)	<0.001	0.758(0.733 ~ 0.783)	<0.001
吃蔬菜(以 <1 次/天为参照)				
≥ 1 次/天	1.027(0.992 ~ 1.065)	0.135	0.861(0.828 ~ 0.895)	<0.001
吃早餐(以从来不吃为参照)				
有时吃	0.843(0.781 ~ 0.910)	<0.001	0.715(0.661 ~ 0.773)	<0.001
天天吃	0.538(0.500 ~ 0.579)	<0.001	0.384(0.356 ~ 0.414)	<0.001
中高强度运动(以 0 天为参照)				
1~3 天	0.888(0.854 ~ 0.923)	<0.001	0.821(0.787 ~ 0.857)	<0.001
≥ 4 天	0.931(0.893 ~ 0.971)	0.001	0.848(0.810 ~ 0.887)	<0.001
体育课时数(以 0~1 节为参照)				
2~3 节	0.826(0.801 ~ 0.853)	<0.001	0.859(0.829 ~ 0.890)	<0.001
≥ 4 节	0.741(0.695 ~ 0.789)	<0.001	0.758(0.705 ~ 0.815)	<0.001
受到严重伤害(以没有为参照)				
有	1.636(1.524 ~ 1.755)	<0.001	1.688(1.569 ~ 1.816)	<0.001
遭受校园欺凌(以没有为参照)				
有	2.315(2.150 ~ 2.492)	<0.001	3.411(3.165 ~ 3.677)	<0.001
被家长打骂(以没有为参照)				
有	2.881(2.766 ~ 3.001)	<0.001	3.191(3.059 ~ 3.328)	<0.001
吸烟(以否为参照)				
是	1.344(1.266 ~ 1.427)	<0.001	1.435(1.349 ~ 1.527)	<0.001
饮酒(以否为参照)				
是	2.062(1.978 ~ 2.151)	<0.001	2.113(2.022 ~ 2.208)	<0.001
网瘾(以否为参照)				
是	3.635(3.289 ~ 4.016)	<0.001	3.976(3.611 ~ 4.377)	<0.001
常量	0.377	<0.001	0.359	<0.001

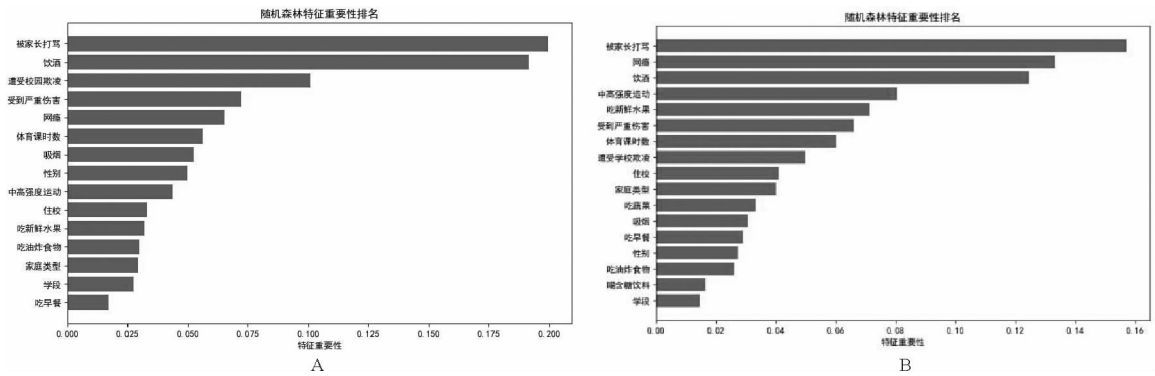
2.4 基于 logistic 回归分析结果构建随机森林模型

2.4.1 自变量重要性排序 利用机器学习的随机化原则, 将 SMOTE 算法处理过的数据按 8:2 比例拆分

为训练集和验证集, 构建随机森林模型, 以变量重要性评分来评价各自变量对因变量的影响程度大小。结果显示, 影响焦虑情绪的前五位因素是被家长打

骂、饮酒、遭受校园欺凌、受到严重伤害和网瘾,影响抑郁情的的前五位因素是被家长打骂、饮酒、网瘾、中高

强度运动和吃新鲜水果,见图 1。



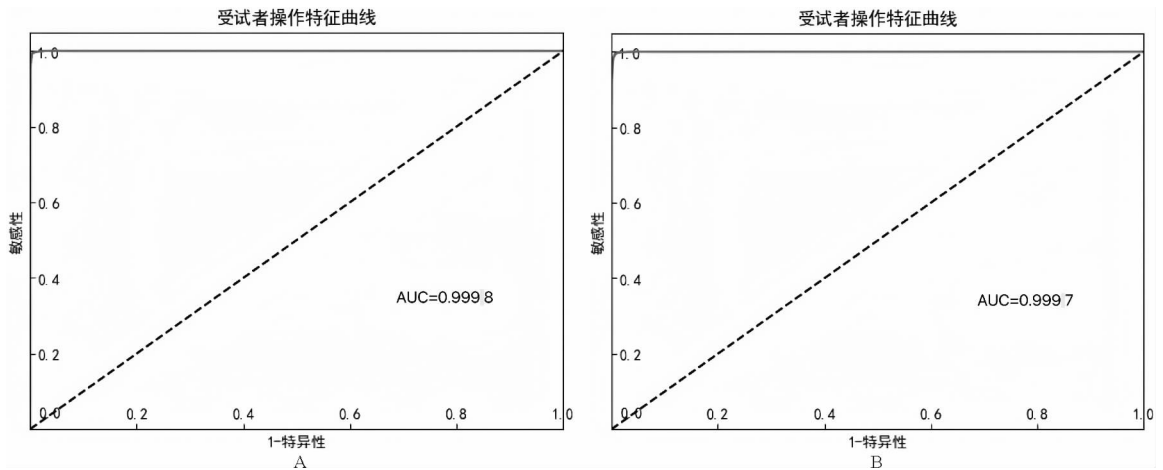
注:(A)焦虑情绪的变量重要性排序;(B)抑郁情绪的变量重要性排序。

图 1 变量重要性排序

Fig. 1 Importance ranking of variables

2.4.2 模型验证 利用随机森林算法在训练集构建中学生焦虑和抑郁情绪发生的风险预测模型并对测试集进行预测,结果显示,焦虑情绪风险预测模型的

灵敏度为 99.53%,特异度为 99.47%,AUC 为 0.999 8;抑郁情绪风险预测模型的灵敏度为 99.67%,特异度为 98.98%,AUC 为 0.999 7,见图 2。



注:(A)预测焦虑情绪风险的 ROC 曲线;(B)预测抑郁情绪风险的 ROC 曲线。

图 2 随机森林算法预测焦虑、抑郁情绪风险的 ROC 曲线

Fig. 2 Random forest algorithm predicts ROC curves for risk of anxiety and depression

3 讨论

一项系统综述显示^[12],全球中学生焦虑检出率为 2.32% ~ 58.79%,抑郁检出率为 2.21% ~ 63.76%;我国中学生焦虑、抑郁检出率分别为 28.0% (95% CI: 20.0 ~ 29.0%)、28.4% (95% CI: 24.4 ~ 32.4%)^[13]。本研究中学生焦虑和抑郁检出率分别为 33.26%、23.72%,检出率差异可能源于地域、经济、学习阶段构成等因素,这也表明山西地区中学生焦虑抑郁现状不容乐观,应及早识别风险因素并干

预。

随机森林广泛应用于学龄儿童、老年人及孕产妇等特殊人群的焦虑抑郁风险预测模型^[14-15],中学生的相关研究较少。本研究构建的中学生焦虑抑郁风险模型 AUC 值接近 1,具有较好的预测效能。但随机森林算法输出的预测概率易受到不平衡数据的影响而造成偏倚。本研究用 SMOTE 方法进行数据处理,将随机森林算法结合多因素 logistic 回归进行直观解释,结果与其他文献报道^[16-18]一致,女生、高中和职高、来自非常态家庭及住校的学生容易出现焦虑抑郁

情绪,如被家长打骂、受到严重伤害、遭受校园欺凌、吸烟、饮酒、网瘾等不良经历是重要预测因素,而健康饮食(吃新鲜水果和蔬菜、吃早餐)则是焦虑抑郁的保护因素。这些因素涵盖社会人口学特征、学校和家庭、健康危险行为等方面,充分表明中学生心理健康促进工作的综合性。

女生可能由于青春早期雌激素水平的变化,社会化的性别角色期望和不平等,更倾向于负面思维模式。高中生面临较高的父母期望和较繁重的学习任务,家庭结构不完整导致安全感和亲子沟通的缺失,在增加焦虑抑郁情绪的同时,还会导致不健康饮食、缺乏体育活动、物质滥用以及欺凌伤害等健康风险行为的发生。有研究^[19]表明长期生活在充斥家庭和校园暴力环境的学生,通过吸烟、饮酒、沉迷网络等以“自我治疗假说”解释的行为来逃避现实生活,这些行为多协同发生,其累积负性水平与焦虑抑郁呈显著剂量反应关系^[20]。此外,饮食行为对于心理健康的影响开始受到关注。食用甜味饮料、油炸食品等西方膳食模式在纵向研究中已被证实与焦虑抑郁风险增加有关^[21],而摄入水果蔬菜、吃早餐等健康的生活方式可能通过减少氧化应激和炎性细胞因子来改变神经递质的浓度,进而减轻不良情绪^[22]。有研究^[18]表明上述健康风险行为具有协同效应,且风险行为和焦虑抑郁之间可能存在双向关系,尚需进一步基于聚类 and 潜在类别分析探讨。

本研究存在一定局限性。第一,问卷中的数据信息多来源于调查对象的主观回忆,可能存在一定的回忆偏倚。第二,本研究属于横断面调查,可反映焦虑、抑郁情绪的危险因素,但不能确定变量间的因果关系,也不能排除研究对象的既往史、家族史等可能的混杂因素。第三,由于研究对象仅为山西地区中学生且未进行外部验证,该模型的推广和应用仍需要增加样本量和前瞻性研究来进一步验证和改进。

4 结 论

综上所述,2023 年山西省中学生焦虑、抑郁情绪的检出率较高,除性别、学段等基本人口学特征外,其健康风险行为与焦虑、抑郁情绪呈正相关,而健康饮食和体育锻炼有利于降低焦虑抑郁情绪的风险。随机森林模型具有较好的预测能力,可获知各个特征的重要性。在公共卫生工作中,可构建专业化的随机森林模型对焦虑和抑郁高危人群进行识别和筛选,并采取相应措施,开展心理健康筛查,早期预测与识别,促进青少年学生心理的健康成长。

志谢 本研究得到山西省疾病预防控制中心的大力支持

利益冲突声明 本研究不存在任何利益冲突

参考文献

- [1] D'agostino A, Gritti ES, Gagliardi C. Generation COVID – 19 and bodily disorders: Hyperbolic narratives and a developmental psychopathology perspective[J]. *Frontiers in Public Health*, 2022, 10: 976584.
- [2] GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990 – 2019; a systematic analysis for the Global Burden of Disease Study 2019[J]. *Lancet*, 2020, 396(10258): 1204 – 1222.
- [3] 徐涛,周县委,张天成,等. 中国中学生抑郁症状危险因素的 Meta 分析[J]. *现代预防医学*, 2022, 49(5): 814 – 818, 844. Xu T, Zhou XW, Zhang TC, et al. Risk factors of depression symptom among Chinese middle school students: a meta – analysis [J]. *Modern Preventive Medicine*, 2022, 49(5): 814 – 818, 844. (In Chinese)
- [4] 汪妍,彭畅,程君茜,等. 中国 5 省在校中学生自杀行为流行状况及其与抑郁、焦虑关联[J]. *中国公共卫生*, 2023, 39(10): 1225 – 1231. Wang Y, Peng C, Cheng JH, et al. Prevalence of and impact of depression and anxiety on suicidal behaviors among high school students: a cross – sectional survey in five provinces, China [J]. *Chinese Journal of Public Health*, 2023, 39(10): 1225 – 1231. (In Chinese)
- [5] 高悦,肖婉,魏瑞虹,等. 童年期不良经历和积极经历与中学生抑郁焦虑症状的关联[J]. *中国学校卫生*, 2024, 45(8): 1075 – 1079. Gao Y, Xiao W, Wei RH, et al. Correlation of adverse and positive childhood experiences and depression and anxiety symptoms among middle school students [J]. *Chinese Journal of School Health*, 2024, 45(8): 1075 – 1079. (In Chinese)
- [6] 张炎圣,杨艳杰,乔正学. 重大突发公共卫生事件下中国青少年焦虑情绪现状及其影响因素分析[J]. *中国公共卫生*, 2023, 39(4): 426 – 429. Zhang YS, Yang YJ, Qiao ZX. Anxiety emotion and its associates among adolescents during a major public health emergency: a cross – sectional survey in China [J]. *Chinese Journal of Public Health*, 2023, 39(4): 426 – 429. (In Chinese)
- [7] 邓淦元,杨中婷,杜兴梅,等. 基于随机森林算法和多因素 logistic 回归分析模型的孕期抑郁影响因素[J]. *中华疾病控制杂志*, 2023, 27(9): 1003 – 1009. Deng GY, Yang ZT, Du XM, et al. Exploring the influencing factors of antenatal depression based on random forest algorithm and multivariate logistic regression model [J]. *Chinese Journal of Disease Control & Prevention*, 2023, 27(9): 1003 – 1009. (In Chinese)
- [8] 中华人民共和国疾病预防控制局. 国家卫生健康委办公厅关于印发 2019 年全国学生常见疾病和健康影响因素监测工作方案的通知[EB/OL]. [2024 – 12 – 11]. <http://www.nhc.gov.cn/jkj/s5898bm/201903/1bcbac21e1864377ad24984fac014c7d.shtml>. Bureau of Disease Control, Prevention and Prevention, PRC. Notice of the General Office of the National Health Commission of the PRC on issuing the work program of monitoring and intervention of common diseases and health influencing factors among students in 2019 [EB/OL]. [2024 – 12 – 11]. <http://www.nhc.gov.cn/jkj/s5898bm/201903/1bcbac21e1864377ad24984fac014c7d.shtml>. (in Chinese)

- 推断研究[J]. 中国循证医学杂志, 2023, 23(6): 621-626.
- Tian J, Du YT, Wang YJ, et al. The causal effects of statins to the prognosis of chronic heart failure[J]. Chinese Journal of Evidence-Based Medicine, 2023, 23(6): 621-626. (In Chinese)
- [13] Sinha P, Calfee CS, Delucchi KL. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls [J]. Critical Care Medicine, 2021, 49(1): e63-e79.
- [14] 马雪祺, 王惠萍, 吕军, 等. 基于有序 logistic 模型的 0~3 岁儿童乘车行为影响因素分析[J]. 中国卫生统计, 2022, 39(3): 389-392, 396.
- Ma XQ, Wang HP, Lv J, et al. Analysis of influencing factors of 0-3 year old children's ride behavior based on ordered logistic model[J]. Chinese Journal of Health Statistics, 2022, 39(3): 389-392, 396. (In Chinese)
- [15] 杨弘, 田晶, 王可, 等. 混合型缺失数据填补方法比较与应用 [J]. 中国卫生统计, 2020, 37(3): 395-399.
- Yang H, Tian J, Wang K, et al. Comparison and application of hybrid missing data filling methods[J]. Chinese Journal of Health Statistics, 2020, 37(3): 395-399. (In Chinese)
- [16] Venkatesan P. Gold COPD report; 2024 update[J]. The Lancet Respiratory Medicine, 2024, 12(1): 15-16.
- [17] Shan J, Yin RL, Panuthai S. Body mass index and multimorbidity risk: A systematic review and dose-response meta-analysis[J]. Archives of Gerontology and Geriatrics, 2024, 123: 105418.
- [18] Powell-Wiley TM, Poirier P, Burke LE, et al. Obesity and cardiovascular disease: a scientific statement from the American heart association [J]. Circulation, 2021, 143(21): e984-e1010.
- [19] Zhang XF, Chen HR, Gu KF, et al. Association of body mass index with risk of chronic obstructive pulmonary disease: a systematic review and Meta-Analysis[J]. COPD, 2021, 18(1): 101-113.
- [20] 关日辉, 孙杰, 吴东, 等. 生理浓度血清胆红素水平对慢性阻塞性肺疾病急性加重的预测价值[J]. 广东医科大学学报, 2020, 38(3): 297-300.
- Guan RH, Sun J, Wu D, et al. Predictive value of physiological concentration of serum bilirubin on acute exacerbation of chronic obstructive pulmonary disease [J]. Journal of Guangdong Medical College, 2020, 38(3): 297-300. (In Chinese)
- [21] Macdonald DM, Kunisaki KM, Wilt TJ, et al. Serum bilirubin and chronic obstructive pulmonary disease (COPD): a systematic review[J]. BMC Pulmonary Medicine, 2021, 21(1): 33.
- [22] Izcovich A, Ragusa MA, Tortosa F, et al. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review[J]. PLOS One, 2020, 15(11): e0241955.
- [23] 陈妙. 呼吸康复护理干预对慢性阻塞性肺病患者运动耐力及呼吸困难指数的影响[J]. 中外医学研究, 2022, 20(11): 109-112.
- Chen M. Effect of respiratory rehabilitation nursing intervention on exercise tolerance and dyspnea index in patients with COPD [J]. Chinese and Foreign Medical Research, 2022, 20(11): 109-112. (In Chinese)
- [24] Hess MW. Oxygenotherapy in COPD [J]. Respiratory Care, 2023, 68(7): 998-1012.

收稿日期: 2024-08-27

(上接第 273 页)

- [9] Spitzer RL, Kroenke K, Williams JB, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7 [J]. Archives of Internal Medicine, 2006, 166(10): 1092-1097.
- [10] Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population [J]. Applied Psychological Measurement, 1977, 1(3): 385-401.
- [11] Xu ZZ, Shen DR, Nie TZ, et al. A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data [J]. Journal of Biomedical Informatics, 2020, 107: 103465.
- [12] Racine N, Mearthur BA, Cooke JE, et al. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: A meta-analysis [J]. JAMA Pediatrics, 2021, 175(11): 1142-1150.
- [13] Zhang YY, Bao XQ, Yan JX, et al. Anxiety and depression in Chinese students during the COVID-19 pandemic: a Meta-Analysis [J]. Frontiers in Public Health, 2021, 9: 697642.
- [14] Qasrawi R, Vicuna polo SP, Abu al-Halawa D, et al. Assessment and prediction of depression and anxiety risk factors in schoolchildren: machine learning techniques performance analysis [J]. JMIR Formative Research, 2022, 6(8): e32736.
- [15] Tabares tabares M, Vélez Álvarez C, Bernal salcedo J, et al. Anxiety in young people: Analysis from a machine learning model [J]. Acta Psychologica, 2024, 248: 104410.
- [16] Tang ZT, Yang XY, Tan WY, et al. Patterns of unhealthy lifestyle and their associations with depressive and anxiety symptoms among Chinese young adults: A latent class analysis [J]. Journal of Affective Disorders, 2024, 352: 267-277.
- [17] Hwang SY, Nam JY. Do lifestyle risk behaviors affect anxiety, depressive symptoms, and suicidal ideation among adolescents? [J]. Psychiatry Research, 2022, 318: 114942.
- [18] Wang M, Mou X, Li T, et al. Association between comorbid anxiety and depression and health risk behaviors among Chinese adolescents: cross-sectional questionnaire study [J]. JMIR Public Health and Surveillance, 2023, 9: e46289.
- [19] Kleppang AL, Vettore MV, Hartz I, et al. Patterns of unhealthy behaviours during adolescence and subsequent anxiety and depression in adulthood: a prospective register linkage study of the HUNT survey and health registries [J]. The International Journal of Behavioral Nutrition and Physical Activity, 2023, 20(1): 6.
- [20] 王娇娇. 累积社会生态风险因素与青少年聚集性健康危害行为关联方向性研究 [D]. 合肥: 安徽医科大学, 2024.
- Wang JJ. The study on the directivity of association between cumulative social-ecological risk factors and clustering health risk behaviors among adolescents [D]. Hefei: Anhui Medical University, 2024. (In Chinese)
- [21] Reales-Moreno M, Tonini P, Escorihuela RM, et al. Ultra-Processed foods and drinks consumption is associated with psychosocial functioning in adolescents [J]. Nutrients, 2022, 14(22): 4831.
- [22] Zhou J, Wang H, Pao C, et al. Association between 29 food groups of diet quality questionnaire and perceived stress in Chinese adults: a prospective study from China health and nutrition survey [J]. BMC Public Health, 2024, 24(1): 1832.

收稿日期: 2024-07-09