

# 基于机器学习的中小学生学习近视影响因素分析及预测模型的开发与验证

顾泉成<sup>1,2</sup>, 陈新丽<sup>3</sup>, 陈 暕<sup>4</sup>, 孟 聪<sup>1</sup>, 段海平<sup>4</sup>

引用: 顾泉成, 陈新丽, 陈暕, 等. 基于机器学习的中小学生学习近视影响因素分析及预测模型的开发与验证. 国际眼科杂志, 2025, 25(2): 328-336.

作者单位:<sup>1</sup>(250012) 中国山东省济南市, 山东大学齐鲁医学院公共卫生学院;<sup>2</sup>(266071) 中国山东省青岛市市南区疾病预防控制中心;<sup>3</sup>(266033) 中国山东省青岛市公共卫生临床中心;<sup>4</sup>(266033) 中国山东省青岛市疾病预防控制中心

作者简介: 顾泉成, 男, 毕业于潍坊医学院, 在读硕士研究生, 副主任医师, 研究方向: 学生常见病防控。

通讯作者: 段海平, 女, 博士, 主任医师, 研究方向: 环境卫生和慢性病防制及近视防控. duan\_hp@126.com

收稿日期: 2024-06-25 修回日期: 2024-12-19

## 摘要

**目的:** 筛选分析中小学生学习近视影响因素并建立预测模型, 为儿童青少年的近视防控措施提供思路。

**方法:** 于 2023-09 以学校为单位, 采用分层整群抽样的方式抽取青岛市城区 2 所小学、2 所初中、2 所高中及 1 所职高中小学生学习 1 759 人, 开展近视筛查和影响因素问卷调查。筛查和判定依据主要按照《标准对数视力表》(GB/T11533-2011)、《儿童青少年近视筛查规范》进行。基于机器学习算法 LASSO 联合 XGBoost 对近视影响因素分析及预测模型开发, 以交互式 Nomogram 实现可视化。采用 R 统计软件版本 4.3.3 进行统计分析。

**结果:** 青岛市城区中小学生学习筛查性近视率为 70.61% (1 242 人)。筛选的最佳预测变量分别为年级、性别、父母是否近视、每天室内静坐时间、读写时眼睛距离书本超过一尺、每天的睡眠时间、看电视/玩电视游戏时眼睛距离电视显示屏的距离超过 3 m、课间活动场所、补习班共多长时间、近距离用眼时多长时间休息一次眼睛、每天用电脑的时间、平均每天放学后做作业时间等 12 项影响因素。预测模型训练集、测试集的 AUC 分别为 0.770 (95% CI: 0.751-0.789)、0.732 (95% CI: 0.714-0.750)。

**结论:** 开发和验证了一个基于机器学习筛选近视影响因素的中小学生学习近视发病风险的预测模型, 并实现了可视化。

**关键词:** 中小学生学习; 近视; 影响因素; 机器学习; 预测模型

DOI: 10.3980/j.issn.1672-5123.2025.2.29

## Development and verification of prediction model for influencing factors of myopia among primary and middle school students based on machine learning

Gu Xiaocheng<sup>1,2</sup>, Chen Xinli<sup>3</sup>, Chen Jian<sup>4</sup>, Meng Cong<sup>1</sup>, Duan Haiping<sup>4</sup>

<sup>1</sup>School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 250012, Shandong Province, China; <sup>2</sup>Center for Disease Control and Prevention of Shinan District, Qingdao 266071, Shandong Province, China; <sup>3</sup>Qingdao Public Health Clinical Center, Qingdao 266033, Shandong Province, China; <sup>4</sup>Qingdao Municipal Center for Disease Control and Prevention, Qingdao 266033, Shandong Province, China

**Correspondence to:** Duan Haiping. Qingdao Municipal Center for Disease Control and Prevention, Qingdao 266033, Shandong Province, China. duan\_hp@126.com

Received: 2024-06-25 Accepted: 2024-12-19

## Abstract

• **AIM:** To screen and analyze the influencing factors of myopia among primary and secondary school students and establish a predictive model to provide ideas for the prevention and control measures of myopia among children and adolescents.

• **METHODS:** A total of 1 759 primary and secondary school students from 2 primary schools, 2 junior high schools, 2 senior high schools and 1 vocational high school in the urban area of Qingdao were sampled by means of stratified cluster sampling in September 2023. Vision screening and a questionnaire survey on influencing factors were carried out based on machine learning algorithms. The screening and determination were mainly conducted in accordance with the Standard Logarithmic Visual Acuity Chart (GB/T11533-2011) and the Specifications for Screening Myopia in Children and Adolescents. The influencing factors of myopia were analyzed and a prediction model was developed based on the machine learning algorithms LASSO in combination with XGBoost, and visualization was achieved through an interactive Nomogram. Statistical analysis was performed using R statistical software version 4.3.3.

• **RESULTS:** The screening prevalence of myopia among

primary and secondary school students in the urban area of Qingdao was 70.61% (1 242 cases). The optimal predictive variables for screening were grade, gender, whether parents were myopic, daily indoor sedentary time, appropriate distance between eyes and books during reading and writing, daily sleep time, distance between eyes and TV screen when watching TV/playing video games exceeding 3 meters, the playground during breaks, total duration of tutorial classes, how often eyes are rested during near work, daily computer usage time, and average daily homework time after school, totaling 12 influencing factors. The AUCs of the training set and test set were 0.770 (95% CI: 0.751–0.789) and 0.732 (95% CI: 0.714–0.750), respectively.

• **CONCLUSION:** A machine learning-based prediction model was developed and validated to predict the risk of myopia onset in primary and secondary school students, accompanied by effective visualization techniques.

• **KEYWORDS:** primary and secondary school students; myopia; influencing factors; machine learning; prediction model

**Citation:** Gu XC, Chen XL, Chen J, et al. Development and verification of prediction model for influencing factors of myopia among primary and middle school students based on machine learning. *Guoji Yanke Zazhi (Int Eye Sci)*, 2025,25(2):328–336.

## 0 引言

近视已成为日益严重的公共卫生问题,给患者带来严重的不便。我国儿童青少年 2020 年总体近视率为 52.7%<sup>[1]</sup>,预计到 2030 年我国将有 61.8% 的儿童青少年近视<sup>[2]</sup>。可见,儿童青少年近视是我国中小学生的常见病之一,防控学生近视是学校卫生的工作重点<sup>[2-3]</sup>。既往研究显示,近视的形成受遗传因素和环境因素共同影响,但成因机制目前还不完全清楚<sup>[4-5]</sup>。目前国内外对学生近视形成的影响因素研究主要通过 Logistic 回归开展<sup>[6]</sup>,本研究希望通过机器学习算法对青岛市城区 2023 年全国学生常见病监测项目的学生调查数据开展中小学生近视的影响因素筛选,并开发和验证一种可视化预测模型,为中小学生近视精准防控提供科学依据。

## 1 对象和方法

**1.1 对象** 本次研究对象为青岛市城区中小學生。于 2023-09 以学校为单位,采用分层整群抽样的方式抽取青岛市城区的 2 所小学、2 所初中、2 所高中及 1 所职高中小學生。以自然班级为最小整群,根据平均约 60 人/班和样本要求,随机从被选择的学校抽取 2-3 个班进行视力检查和问卷调查。共纳入学生 1 759 人,其中小学生 509 人 (28.94%)、初中生 475 人 (27.00%)、高中生 527 人 (29.96%)、职高生 248 人 (14.10%);男生 961 人 (54.63%)、女生 798 人 (45.37%)。对于机器学习来说,样本量越大,越能增强其构建的预测模型的鲁棒性,1 759 人高于样本量计算器<sup>[7]</sup>估值 528 人,也符合 10 EPV (10 events per

variable)原则。纳入标准:研究对象的年级阶段为 4-12 年级,能够配合视力筛查且能独立完成问卷调查;排除标准:可能干扰视力监测结果的已知的严重健康问题。本研究严格遵守一般伦理学原则(批准号:2024101),在获得参与者及其监护人知情同意后开展调查。

## 1.2 方法

**1.2.1 问卷调查** 采用 2023 年全国中小學生常见病及健康影响因素监测方案中《学生视力不良、脊柱弯曲异常及影响因素专项调查表》,并依托青岛市学生健康监测平台实施线上无纸化调查。问卷内容涉及学生基本信息、校内用眼环境、校外用眼情况、读写姿势、电子屏幕使用情况、近距离用眼习惯、户外活动及睡眠情况、近视检查及矫治情况等 8 大项 37 小项近视影响因素,见表 1。本次研究选取视力标准判定任 1 眼裸眼视力 < 5.0 或戴角膜接触镜的学生为筛查性近视<sup>[8-9]</sup>。

**1.2.2 质量控制** 由经过统一培训的卫生老师向调查对象发放在线问卷,并解释问卷内容和要求。安排督导人员进行现场质控,完成问卷后,查缺补漏以确保正确率。

**统计学方法:**采用 R 统计软件版本 4.3.3 进行统计分析。分类资料采用频数或百分比表示,组间比较采用卡方检验或 Fisher 检验。非正态分布定量资料用  $M(P_{25}, P_{75})$  表示,两样本比较使用 Mann-Whitney *U* 检验。机器学习算法与传统统计方法相比,对数据的限制更少,并且具有有效建模复杂数据集的能力。使用 LASSO 联合 XGBoost 算法对中小學生近视影响因素进行筛选,以训练集、测试集 8:2 进行近视预测模型构建和验证,以交互式 Nomogram 实现预测模型的可视化。主要用到 glmnet、xgboost、ggplot2、pROC 等程序包。所有统计检验均为双侧检验,检验水准  $\alpha=0.05$ 。

## 2 结果

**2.1 基本情况** 青岛市城区中小學生筛查性近视率为 70.61% (1 242 人),其中男生检出率 66.91% (643 人),女生检出率 75.06% (599 人),差异有统计学意义(均  $P < 0.01$ )。四至六年级的近视检出率分别为:40.68%、52.73%、53.29%;初一至初三近视检出率分别为:76.67%、76.07%、79.01%;高一至高二的近视检出率分别为:82.29%、84.18%、84.81%;职高一至职高二的近视检出率分别为 73.17%、75.58%、76.25%。年级间总体比较差异有统计学意义( $\chi^2=182.669, P < 0.001$ ),其中四、五年级组间比较差异有统计学意义( $\chi^2=4.98, P < 0.05$ );四、六年级组间比较差异有统计学意义( $\chi^2=5.49, P < 0.05$ );五、六年级进行组间比较,差异无统计学意义( $\chi^2=0.01, P=0.92$ )。此外,近视检出率在是否住校、民族、每天做几次眼保健操等影响因素差异均有统计学意义(均  $P < 0.01$ ),见表 2。年级、性别、课间休息时活动场所、眼睛距离书本超过一尺、做作业时间、近距离用眼时多长时间休息一次眼睛等与近视的关系见图 1。

**2.2 中小學生近视影响因素(预测变量)的筛选** LASSO 回归将分类变量进行哑变量设置,进行 10 折交叉验证,在  $\lambda_{min}=0.01$  时取得系数非零变量。LASSO 系数分布图、交叉验证图分别见图 2、3。XGBoost 进行了模型拟合,并按

表 1 变量对应表

变量	影响因素
A01	年级
A02	性别
A03	是否住校
A04	汉族
B01	是否定时调换座位
B02	课桌椅高度是否会根据你的身高进行调整
B03	每天做几次眼保健操
B04	课间活动场所
C01	平均每天放学后做作业时间
C02	补习班共多长时间
C03	从几岁开始参加课外学习
C04	上小学前参加过哪类课外学习班
C05	家长会减少你运动的时间吗
C06	家长是否限制你看电视、玩电脑或电子游戏时间
D01	读写时,胸口离桌子边沿超过一拳
D02	读写时,眼睛距离书本超过一尺
D03	读写时,手指距离笔尖一寸左右
D04	老师是否提醒你注意读写姿势
D05	父母是否提醒你注意读写姿势
E01	你平均每天看多长时间电视
E02	通常你每天用电脑的时间
E03	你平均每天使用移动电子设备(包括手机、掌上游戏机、平板电脑等)时间
E04	你平均每天看电视、使用电脑、手机、游戏机、以及看(玩)其他电子屏幕的时间
F01	你在阳光直射下看书或电子屏幕
F02	你在天黑后看电子屏幕时关灯吗
F03	你会躺着或趴着看书或电子屏幕吗
F04	你在走路或乘车时看书或电子屏幕吗
F05	你天黑后在家读书写字用什么灯光
F06	你在用电脑时,眼睛距离电脑显示屏的距离超过 66 cm
F07	你在看电视/玩电视游戏时,眼睛距离电视显示屏的距离超过 3 m 吗
F08	在近距离用时,多长时间休息一次眼睛
G01	过去 1 wk 里,你每天白天户外活动时间是多少
G02	平均每天的睡眠时间
G03	你每天在室内静坐的时间是多少
H01	你的父母是否近视
H02	你过去 1 a 内做过几次视力检查
H03	不戴眼镜的情况下,你是否有一只或两只眼睛的视力低于 5.0

表 2 影响因素的组间比较

影响因素	类别	正常组(n=517)	近视组(n=1242)	$\chi^2$ /Fisher/Z	P
年级(人,%)	四年级	105(20.31)	72(5.80)	182.669	<0.001
	五年级	78(15.09)	87(7.00)		
	六年级	78(15.09)	89(7.16)		
	初一	35(6.77)	115(9.26)		
	初二	39(7.54)	124(9.98)		
	初三	34(6.58)	128(10.30)		
	高一	34(6.58)	158(12.72)		
	高二	28(5.42)	149(12.00)		
	高三	24(4.64)	134(10.79)		
	职高一	22(4.25)	60(4.83)		
	职高二	21(4.06)	65(5.23)		
	职高三	19(3.67)	61(4.91)		

续表 2 影响因素的组间比较

影响因素	类别	正常组( <i>n</i> =517)	近视组( <i>n</i> =1242)	$\chi^2$ /Fisher/ <i>Z</i>	<i>P</i>
性别(人,%)	男	318(61.51)	643(51.77)	13.575	<0.001
	女	199(38.49)	599(48.23)		
住校(人,%)	否	448(86.65)	951(76.57)	22.188	<0.001
	是	69(13.35)	291(23.43)		
汉族(人,%)	否	21(4.06)	27(2.17)	4.217	0.040
	是	496(95.94)	1215(97.83)		
每天眼保健操(人,%)	1次	6(1.16)	47(3.78)	11.823	<0.001
	2次	466(90.14)	1109(89.29)		
	3次及以上	44(8.51)	70(5.64)		
	在校不做	1(0.19)	16(1.29)		
课间活动场所(人,%)	教学楼内	163(31.53)	668(53.78)	71.659	<0.001
	户外操场	354(68.47)	574(46.22)		
每天放学后做作业时间(人,%)	不到1h	181(35.01)	272(21.90)	11.823	<0.001
	1-2h	191(36.94)	475(38.24)		
	2-3h	75(14.51)	268(21.58)		
	3h及以上	58(11.22)	206(16.59)		
	未知	12(2.32)	21(1.69)		
补习班时间(人,%)	不到1h/d	66(12.76)	109(8.78)	11.823	0.037
	1-2h/d	112(21.66)	257(20.69)		
	2-3h/d	61(11.80)	180(14.49)		
	3h/d及以上	69(13.35)	177(14.25)		
	未知	11(2.13)	13(1.05)		
从几岁开始参加课外学习(人,%)	无补习班	198(38.30)	506(40.74)	29.191	<0.001
	3岁以前	10(1.93)	28(2.25)		
	2-3岁	31(6.00)	54(4.35)		
	3-4岁	65(12.57)	86(6.92)		
	4-5岁	76(14.70)	132(10.63)		
	5-6岁及以后	238(46.03)	709(57.08)		
家长限制看电视等时间(人,%)	从没参加	97(18.76)	233(18.76)	10.749	0.001
	否	147(28.43)	456(36.71)		
胸口离桌子边沿超过一拳(人,%)	是	370(71.57)	786(63.28)	28.316	<0.001
	从不	41(7.93)	94(7.57)		
	偶尔	135(26.11)	379(30.52)		
	经常	125(24.18)	405(32.61)		
	总是	216(41.78)	364(29.31)		
眼睛距离书本超过一尺(人,%)	从不	32(6.19)	92(7.41)	48.426	<0.001
	偶尔	127(24.56)	375(30.19)		
	经常	126(24.37)	428(34.46)		
	总是	232(44.87)	347(27.94)		
手指距离笔尖一寸(人,%)	从不	29(5.61)	97(7.81)	11.627	0.009
	偶尔	121(23.40)	319(25.68)		
	经常	131(25.34)	363(29.23)		
	总是	236(45.65)	463(37.28)		
老师提醒注意读写姿势(人,%)	从不	44(8.51)	141(11.35)	11.23	0.011
	偶尔	159(30.75)	416(33.49)		
	经常	127(24.56)	331(26.65)		
	总是	187(36.17)	354(28.50)		
父母提醒注意读写姿势(人,%)	从不	29(5.61)	101(8.13)	14.814	0.002
	偶尔	136(26.30)	321(25.84)		
	经常	125(24.18)	378(30.43)		
	总是	227(43.91)	442(35.59)		

续表 2 影响因素的组间比较

影响因素	类别	正常组 (n=517)	近视组 (n=1242)	$\chi^2$ /Fisher/Z	P
每天用电脑的时间(人,%)	没用过	232(44.87)	519(41.79)	5.228	0.389
	不到 1 h	167(32.30)	442(35.59)		
	1-2 h	75(14.51)	157(12.64)		
	2-3 h	21(4.06)	72(5.80)		
	3-4 h	8(1.55)	22(1.77)		
	4 h 及以上	14(2.71)	30(2.42)		
每天用移动电子设备时间[M(P <sub>25</sub> ,P <sub>75</sub> ),h]		0.5(0,1.00)	0.5(0,1.30)	293736.500	0.004
每天看电子屏幕时间[M(P <sub>25</sub> ,P <sub>75</sub> ),h]		0.3(0.0,1.00)	0.5(0,2.00)	291088.000	0.002
阳光下看书或电子屏幕(人,%)	从不	345(66.73)	687(55.31)	24.055	<0.001
	偶尔	139(26.88)	481(38.73)		
	经常	20(3.87)	53(4.27)		
	总是	13(2.51)	21(1.69)		
	天黑后看电子屏幕时关灯(人,%)	从不	367(70.99)		
	偶尔	88(17.02)	374(30.11)		
	经常	31(6.00)	88(7.08)		
	总是	31(6.00)	59(4.75)		
	躺着趴着看书或电子屏幕(人,%)	从不	301(58.22)		
	偶尔	148(28.63)	515(41.46)		
	经常	51(9.86)	183(14.73)		
	总是	17(3.29)	44(3.54)		
	走路乘车时看书电子屏幕(人,%)	从不	343(66.34)		
	偶尔	132(25.53)	448(36.07)		
	经常	30(5.80)	119(9.58)		
	总是	12(2.32)	26(2.09)		
	眼距电脑超过 66 cm(人,%)	从不	47(9.09)		
	偶尔	89(17.21)	244(19.64)		
	经常	104(20.12)	319(25.68)		
	总是	155(29.98)	359(28.90)		
	不用电脑	122(23.60)	205(16.50)		
眼距电视超过 3 m(人,%)	从不	42(8.12)	89(7.16)	14.339	0.006
	偶尔	69(13.35)	215(17.31)		
	经常	112(21.66)	274(22.06)		
	总是	233(45.07)	463(37.28)		
	不看电视	61(11.80)	201(16.18)		
近距离用眼时多长时间休息一次眼睛(人,%)	15 min 及以下	119(23.02)	188(15.14)	26.141	<0.001
	15 min-0.5 h	120(23.21)	235(18.92)		
	0.5-1 h	132(25.53)	390(31.40)		
	1-2 h	92(17.79)	255(20.53)		
	2-3 h	29(5.61)	82(6.60)		
	3 h 及以上	25(4.84)	92(7.41)		
每天睡眠时间[M(P <sub>25</sub> ,P <sub>75</sub> ),h]		9(7.50,10.00)	8(7.00,9.00)	417836.5	<0.001
每天室内静坐时间[M(P <sub>25</sub> ,P <sub>75</sub> ),h]		5(1.00,9.00)	8(4.00,10.50)	242577	<0.001
父母是否近视(人,%)	只有父亲近视	83(16.05)	211(16.99)	47.452	<0.001
	只有母亲近视	103(19.92)	256(20.61)		
	父母都近视	75(14.51)	344(27.70)		
	父母都不近视	256(49.52)	431(34.70)		

注:近视组:不戴眼镜的情况下,任 1 眼裸眼视力<5.0 或配戴角膜接触镜;正常组:裸眼视力≥5.0。



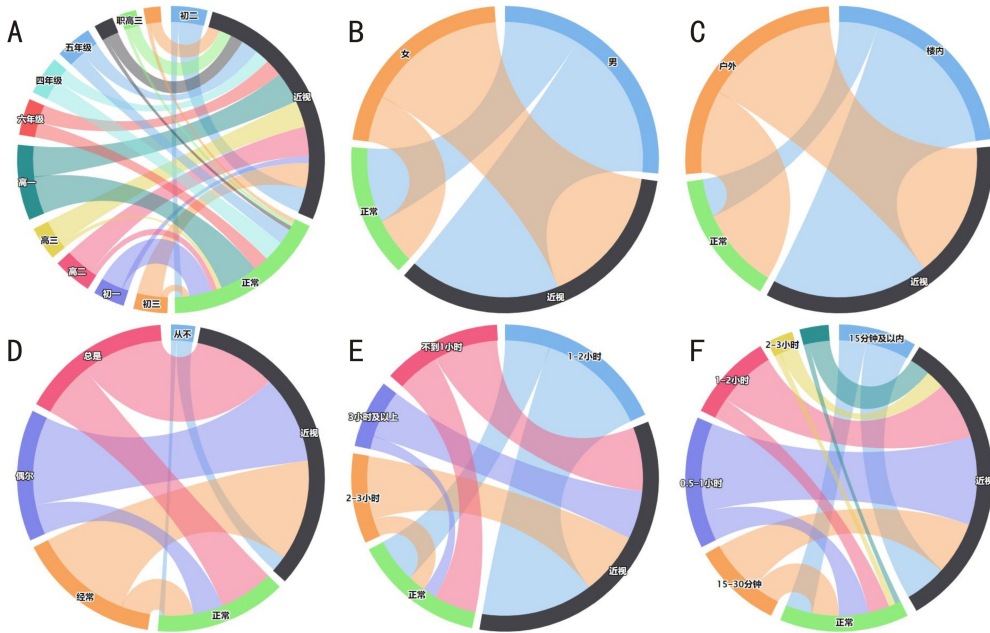


图 1 主要影响因素与近视的关系 A:各学段近视分布;B:性别间近视分布;C:课间活动选择户外楼内对近视影响;D:能否做到眼睛距离书本超过一尺对近视的影响;E:做作业时间对近视的影响;F:近距离用眼多长时间休息一次眼睛对近视的影响。

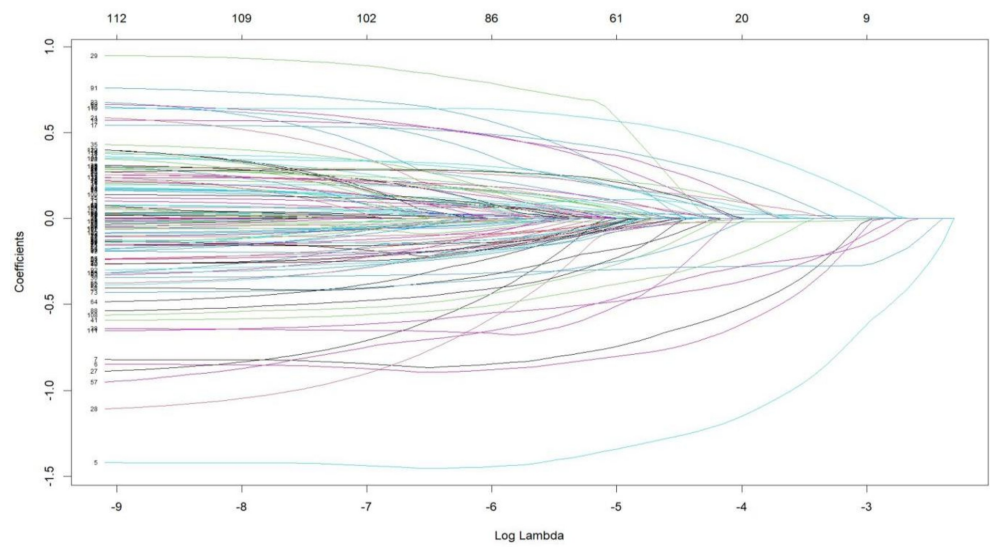


图 2 LASSO 系数分布图。

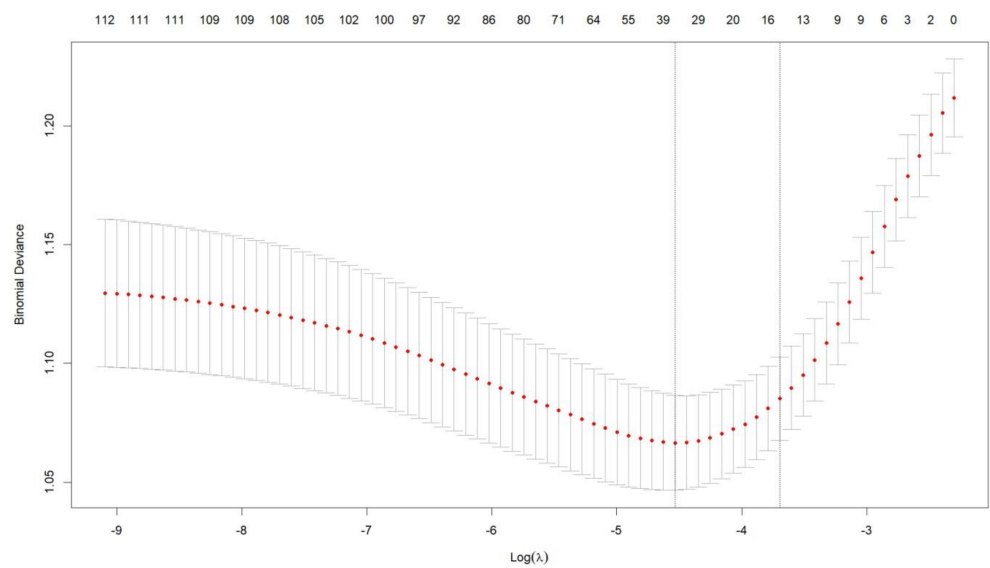


图 3 LASSO 交叉验证。

变量重要性列出系数非零变量,见图4。取二者交集,12项影响因素即:年级、父母是否近视、性别、每天室内静坐时间、读写时眼睛距离书本超过一尺、每天的睡眠时间、看电视/玩电视游戏时眼睛距离电视显示屏的距离超过3 m、课间活动场所、补习班共多长时间、近距离用眼多长时间休息一次眼睛、每天用电脑的时间、平均每天放学后做作业时间。单因素分析显示,高一学生较四年级( $OR=6.78, P<0.05$ )、父母都近视的孩子( $OR=1.8, P<0.05$ )、女生较男生( $OR=1.49, P<0.05$ )、每天放学后做作业时间1-2 h( $OR=1.65, P<0.05$ )及补习班2-3 h( $OR=1.79, P<0.05$ )患近视的风险均为危险因素。而课间在操场活动较室内( $OR=0.4, P<0.05$ )、经常做到眼睛距离书本超过一尺( $OR=0.52, P<0.05$ )、每天的睡眠时间增加( $OR=0.82, P<0.05$ )均为预防近视的保护因素。

**2.3 模型构建与验证** 以上述12项影响因素为预测变量,以是否近视为二分类结局变量。按照训练集、测试

集8:2进行数据划分构建预测模型。训练集、测试集的AUC分别为0.770(95%CI=0.751-0.789)、0.732(95%CI=0.714-0.750)(训练集:用于构建和训练机器学习模型的数据集,模型通过学习训练集中的数据来识别模式和规律,从而能够预测新数据;测试集:用于评估模型在未见过的数据上的性能,这确保了模型能够在真实世界的应用中表现良好)。模型的ROC曲线显示学生近视患病的风险截断值为0.768,模型的ROC、DCA曲线见图5、6。

**2.4 交互式 Nomogram 绘制** 每个预测变量的取值对应列线图第一行的单项得分,然后将12个预测变量的得分相加得到总分,总得分向下对应列线图最后一行“学生近视概率”。总得分越高,说明该学生近视患病的风险也越大。图中显示的是数据集的第一行高三女生近视患病的预测概率为0.861,大于学生近视患病的风险截断值0.768,风险较高,图7。

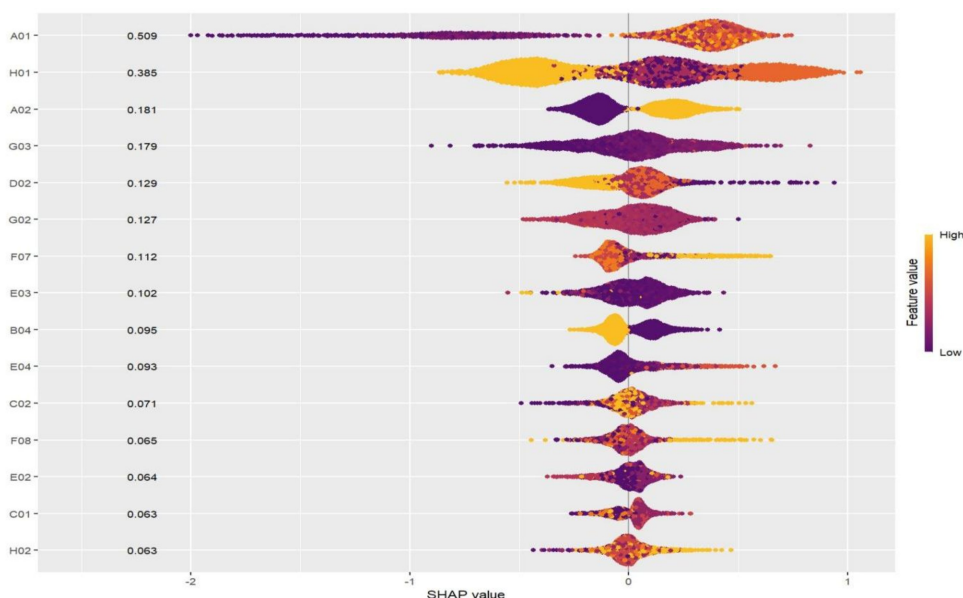


图4 XGBoost 变量筛选。

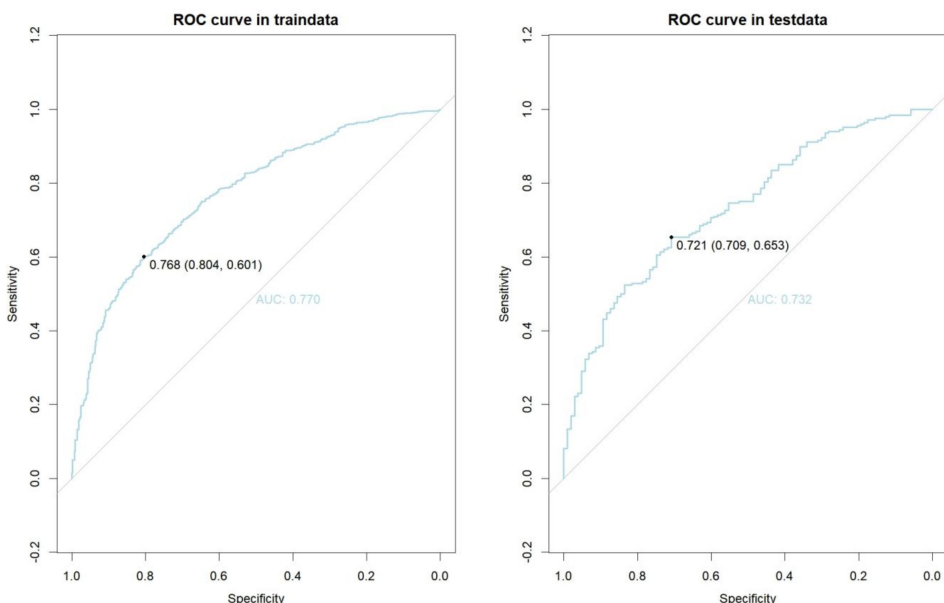


图5 训练集和测试集的ROC曲线 A:训练集的ROC曲线及预测模型截断值;B:测试集的ROC曲线。

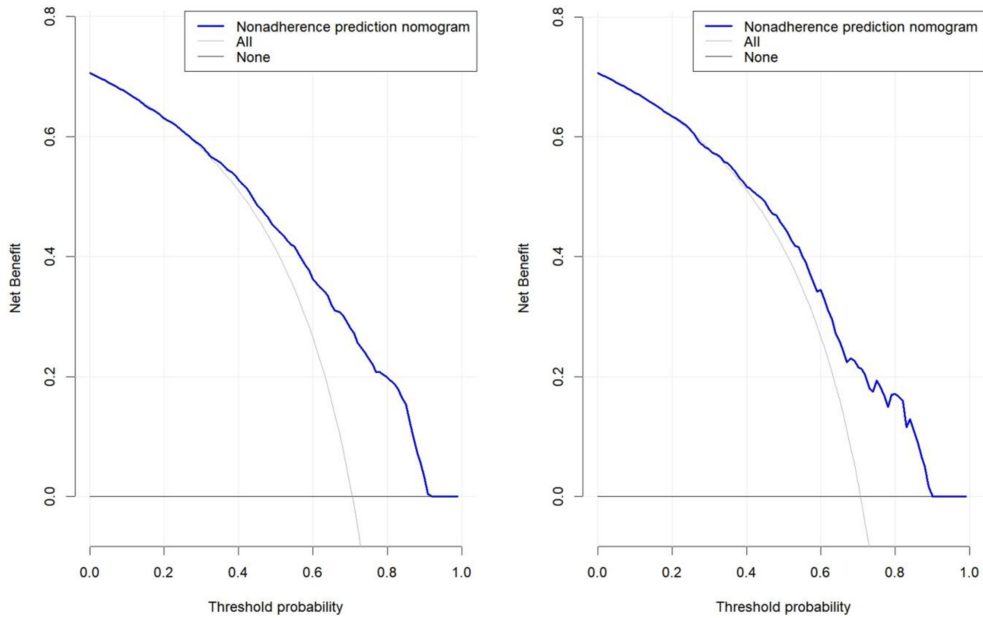


图6 训练集和测试集的DCA曲线 A:训练集构建预测模型的DCA曲线;B:测试集的DCA曲线。

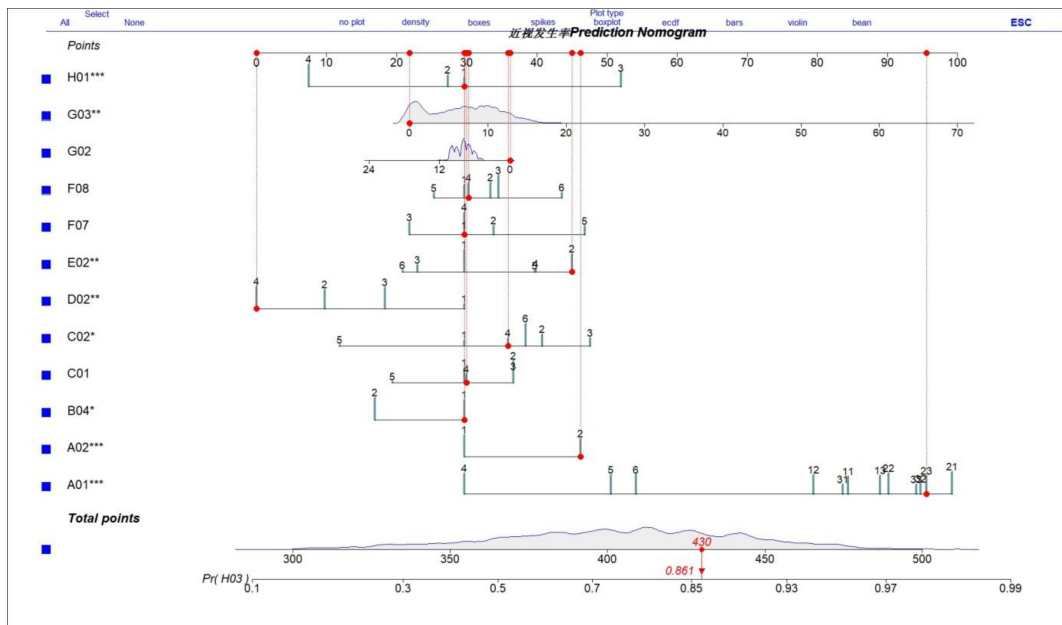


图7 近视发生率的列线图。

### 3 讨论

本次调查结果显示青岛市城区中小學生筛查性近视率70.61%，与云南省(68.3%)、陕西省(69.12%)中小學生近视率相当<sup>[10-11]</sup>，高于2020年全国儿童青少年总体近视率52.7%的调查结果。除五年级、六年级、初一及初三的近视率略高于国家公布的小学平均近视率45.7%、初中生平均近视率76.2%外，其他各年级近视率水平均不高于国家同类水平，且高中生的近视率低于国家高中生平均近视率87.4%的平均水平<sup>[12]</sup>。较高的总体筛查性近视率可能与新冠疫情期間中小學生网课频率增加导致用眼时间延长有关<sup>[10,13]</sup>。另一方面筛查性近视是通过远视力筛查且非睫状肌麻痹状态下电脑验光或串镜检查等方法判定的近视，不等于临床诊断近视，只能筛查出可疑近视学生<sup>[8]</sup>。

本研究结果提示女生近视患病率高于男生，与既往报道一致<sup>[14-16]</sup>。这可能与女生户外活动时间较少，相对文静<sup>[17]</sup>，花更多时间学习，导致用眼过度有关。也可能与女

生青春期较早，眼球发育相对较早且易受外界影响有关。随着年级的升高，筛查性近视率呈现增高趋势，高中阶段近视率达到高峰。这可能与随着年级的升高带来的学习内容、任务不断增加，接触电子产品机会增多，用眼时间增长而导致用眼过多有关，这与国内刘念等<sup>[18]</sup>、陈思明等<sup>[19]</sup>报道一致。职业高中近视率低于同年级普通高中，进一步说明学生近视的发生发展与学业负担及视近时间等有关<sup>[20]</sup>。与2018年的全国小学生筛查性近视率36.0%相比，近视呈现低龄化趋势，尤其是4年级到5年级的近视率上升非常明显，是近视预防的重点<sup>[8]</sup>，要及早干预，防止近视程度的加深及并发症的出现<sup>[21]</sup>。同时也应加强对高中生的近视关注度，虽然高中生的近视风险较低，但其近视程度相对较重，应引起重视<sup>[22]</sup>。

本研究结果显示，课间休息时活动场所选择楼内的近视患病风险高于户外的风险，提示户外活动有助于降低近视发病风险。动物实验和人群流行病学研究均证实了户



外活动对视力的保护作用<sup>[23]</sup>。有研究认为,户外活动预防近视发生的机制与户外运动时视网膜接受到阳光后刺激机体多巴胺分泌增多有关。多巴胺可以有效抑制眼球的拉长,预防近视的发生<sup>[24]</sup>。每天放学后做作业1-2 h较不到1 h ( $OR=1.65, P<0.05$ )、补习2-3 h较不到1 h ( $OR=1.79, P<0.05$ )增加近视患病风险。可能因为这些因素挤占了学生活动时间,导致运动量减少而用眼时间增加所致。长时间高强度的用眼,可导致神经中枢功能衰减和能量代谢代偿异常,产生不可逆的损伤。经常做到眼睛距离书本超过一尺 ( $OR=0.52, P<0.05$ )、每天的睡眠时间增加 ( $OR=0.82, P<0.05$ )减少近视患病风险。近距离用眼是近视发生发展的重要危险因素<sup>[25]</sup>。减少近距离用眼工作,增加远眺的频次可有效降低近视的发生风险<sup>[26]</sup>。睡眠时间增加可相应缩减长时间近距离用眼的机会和时间,不会产生视觉疲劳<sup>[15]</sup>。

本研究发现,父母近视的孩子患近视的风险较高,而且呈现出父母都近视的风险大于仅父亲近视的风险 ( $OR=1.8, P<0.05$ )。如果父母近视,其子女近视风险增加,但这不能完全归于父母遗传基因的传递效应,还可能存在遗传养育效应,即父母的行为和养育方式增加子女近视风险,环境因素成为遗传和近视关联的中介<sup>[27]</sup>。

既往关于学生近视形成的影响因素研究<sup>[28-29]</sup>主要通过 Logistic 回归进行影响因素筛选,且未建立实用的可视化的预测模型。Logistic 对异常值敏感,难以处理多重共线性数据。与之相比,机器学习更容易处理异质性的多维的数据,开发的模型提供更好的预测性能<sup>[30]</sup>。本研究优势在于通过 LASSO 联合 XGBoost 算法对影响因素进行筛选,以交互式 Nomogram 实现预测的可视化,为早期发现易感人群并开展有针对性的近视精准干预提供参考依据。研究的创新点在于开发了一个基于机器学习算法的实用的近视预测模型,具有较较强的临床应用价值。研究提示要重点对具有近视家族史、较低年级学生、女生几类高危人群开展学生视力检查,发现异常及早采取干预措施。本研究也存在一定的局限性,对于室外活动未考虑活动的强度对近视发生发展的影响,需要进一步研究。

综上所述,本文基于机器学习分析中小学生学习近视影响因素并建立验证了一个可以预测中小学习筛查性近视的可视化预测模型。今后将使用前瞻性大样本数据对模型进一步验证和优化,以提高模型的鲁棒性和泛化能力。

**利益冲突声明:** 本文不存在利益冲突。

**作者贡献声明:** 顾泉成论文选题与修改,初稿撰写;陈新丽、陈曛、孟聪文献检索,数据分析;段海平选题指导,论文修改及审阅。所有作者阅读并同意最终的文本。

#### 参考文献

[1] 中华人民共和国国家卫生健康委员会. 国家卫生健康委员会2021年7月13日新闻发布会文字实录. <http://www.nhc.gov.cn/xwzb/webcontroller.do?titleSeq=11389&gectype=1>.

[2] 陈军,何鲜桂,王菁菁,等. 2021至2030年我国6~18岁学生近视眼患病率预测分析. 中华眼科杂志, 2021,57(4):261-267.

[3] 陶芳标. 中国儿童青少年近视病因模型及其政策与策略导向的预防控制. 安徽预防医学杂志, 2022,28(4):261-265.

[4] 郭振,谢森,杜显丽,等. 中国七省份小学生筛查性近视影响因素分析. 中国学校卫生, 2020,41(12):1872-1875.

[5] 李婷,王梅,张东红,等. 沈阳市12~18岁青少年近视相关因素

的探讨及 Nomogram 预测模型的建立. 中国卫生统计, 2022,39(2):238-242.

[6] Shi Y, Zhang GM, Ma CY, et al. Machine learning algorithms to predict intraoperative hemorrhage in surgical patients: a modeling study of real-world data in Shanghai, China. BMC Med Inform Decis Mak, 2023,23(1):156.

[7] Wang X, Ji X. Sample size estimation in clinical research: from randomized controlled trials to observational studies. Chest, 2020,158(1S):S12-S20.

[8] 陶芳标.《儿童青少年近视防控适宜技术指南》专题解读. 中国学校卫生, 2020,41(2):166-168,172.

[9] 国家卫生健康委.《儿童青少年近视防控适宜技术指南》[EB/OL].[2019-10-14]. <http://www.nhc.gov.cn/jkj/s5898bm/201910/c475e0bd2de444379402f157523f03fe.shtml>.

[10] 张金娇,常利涛,肖洁,等. 云南省儿童青少年近视危险因素列线图判别分析. 中国学校卫生, 2023,44(9):1387-1391.

[11] 王妍昕,史伟,王甜,等. 陕西省中小学生学习近视及影响因素研究. 中国慢性病预防与控制, 2023,31(9):651-654.

[12] 近视防控与诊治教育部工程研究中心. 从2019年底到2020年6月,被调查学生近视率增长了11.7%—共同呵护好孩子的眼睛[EB/OL].[2020-10-24]. <http://www.myopcc.com/newsShow.aspx?id=404>.

[13] 谭业丰,谭晓东. 武汉市中小学生学习近视流行现状及影响因素分析. 健康教育与健康促进, 2023,18(5):464-466,470.

[14] 徐文红,王丽娟. 视力障碍青少年身体活动水平和久坐时间调查. 中国学校卫生, 2020,41(2):190-193.

[15] 李红飞,莫健. 中国青少年近视情况及影响因素分析. 现代预防医学, 2021,48(14):2552-2557.

[16] Jones-Jordan LA, Sinnott LT, Chu RH, et al. Myopia progression as a function of sex, age, and ethnicity. Invest Ophthalmol Vis Sci, 2021,62(10):36.

[17] 宋伟,任婉娜,魏兴民,等. 甘南地区儿童青少年近视筛查与矫正情况分析. 国际眼科杂志, 2024,24(2):320-323.

[18] 刘念,何海艳,许云峰,等. 绵阳市中小学生学习视力不良检出率及影响因素分析. 预防医学情报杂志, 2020,36(10):1300-1302,1310.

[19] 陈思明,阮慧蒙,李波,等. 宁波市海曙区儿童和青少年近视及视力情况调查分析. 现代实用医学, 2020,32(11):1363-1364.

[20] 陈黎黎,吴岩,石荣兴,等. 2018—2020年北京市丰台区儿童青少年近视筛查结果分析. 中国健康教育, 2021,37(6):507-510.

[21] 石龙华,荣爽,程茅伟,等. 湖北省中小学生学习近视流行现状及影响因素分析. 现代预防医学, 2021,48(4):649-653.

[22] 高青,刘懿卿,叶茜雯,等. 辽宁省学生近视情况及其影响因素. 中华疾病控制杂志, 2021,25(2):222-226.

[23] 刘忠慧,徐渴,侯常春. 户外活动对儿童近视的影响及机制的研究进展. 中国慢性病预防与控制, 2022,30(7):537-540.

[24] 王硕,吴立娟,刘丽娟,等. 北京市初中一年级学生高度近视现状及影响因素分析. 现代预防医学, 2019,46(14):2567-2571.

[25] 黄坤,李秀红. 青少年近视的影响因素研究进展. 预防医学, 2020,32(6):578-582.

[26] 史浩楠,马少伟,张羽珊,等. 近视的影响因素研究进展. 医学综述, 2021,27(12):2391-2396.

[27] 陶芳标.《儿童青少年近视防控公共卫生综合干预技术指南》专题解读. 中国学校卫生, 2023,44(10):1445-1449.

[28] 谢晓露,吴纲跃,李小英. 金华市婺城区3014名小学学生视力不良现状及影响因素. 国际眼科杂志, 2024,24(2):324-327.

[29] 田向杰,李江,张静,等. 2019年蒙自市中小学生学习近视影响因素分析. 国际眼科杂志, 2022,22(4):652-656.

[30] Ota R, Yamashita F. Application of machine learning techniques to the analysis and prediction of drug pharmacokinetics. J Control Release, 2022,352:961-969.