

决策树和回归技术在中小學生视力影响因素分析中的应用

谢小莲, 李娟, 刘尚红, 孙丽姣, 赵海萍

宁夏医科大学公共卫生与管理学院, 银川 750004

【摘要】 目的 采用决策树模型和 Logistic 回归模型了解宁夏中小學生视力状况及其影响因素, 为制订视力不良防控方案提供依据。**方法** 采用分层整群抽样的方法, 从宁夏银川和吴忠分层整群随机抽取 7~18 岁儿童 3 257 名作为研究对象, 对学生进行了视力检查和问卷调查。应用 Clementine 12.0 中的 C 5.0 算法和 Logistic 回归分析视力影响因素, 利用灵敏度、特异度及正确分类百分比评价 2 个模型的效果。**结果** 宁夏中小學生视力不良总体检出率为 62.6%, 男生、女生视力不良检出率分别为 56.8%, 68.1%, 差异有统计学意义 ($P < 0.01$); 不同学段学生视力不良检出率差异有统计学意义 ($P < 0.01$), 其中高中生检出率最高, 达 82.2%; 决策树 C 5.0 算法分析结果显示, 学段分组、性别、每日做作业的时间及每周进行中等强度锻炼的频率是中小學生视力不良最主要的影响因素; Logistic 回归结果显示, 性别女 ($OR = 1.65, 95\% CI = 1.40 \sim 1.96$)、眼睛距计算机显示屏距离 < 66 cm ($OR = 1.24, 95\% CI = 1.00 \sim 1.52$)、读书时眼睛距离课本不足 1 尺 ($OR = 2.05, 95\% CI = 1.23 \sim 3.40$) 与视力不良呈正相关; 低学段组和放学后做作业时间 < 2 h/d ($OR = 0.76, 95\% CI = 0.62 \sim 0.92$) 与视力不良呈负相关。**结论** 学段、性别和放学后作业时间是影响宁夏中小學生视力状况的主要因素。建议近视预防从低学段组入手, 尽可能减少中小學生视近和课后作业时间, 适量增加尤其是女生的户外活动时间, 从而预防近视的发生和发展。

【关键词】 决策树; 回归分析; 视力; 低; 患病率; 学生

【中图分类号】 R 179 R 778.1 **【文献标识码】** A **【文章编号】** 1000-9817(2019)04-0572-04

Application of decision tree and regression technique in the analysis of visual acuity influencing factors in primary and middle school students/XIE Xiaolian, LI Juan, LIU Shanghong, SUN Lijiao, ZHAO Haiping. School of Public Health and Management, Ningxia Medical University, Yinchuan(750004), China

【Abstract】 Objective To investigate the visual acuity of primary and middle school students in Ningxia by using decision tree model and logistic regression, and to provide evidence for the prevention and control of visual acuity. **Methods** Using stratified cluster sampling method, 3 257 children aged 7-18 years were randomly selected from Yinchuan and Wuzhong in Ningxia. Visual acuity test and questionnaire survey were conducted. The C 5.0 algorithm in Clementine12.0 and Logistic regression were used to analyze the factors associated with visual acuity. **Results** The overall rate of low vision of the middle school of Ningxia was 62.6%, with girls (68.1%) higher than that of boys (56.8%) ($P < 0.01$). There was a significant difference in the prevalence of low vision among different educational stages ($P < 0.01$), with high school students had the highest detection rate (82.2%). Results of decision tree C 5.0 algorithm showed that educational stage, gender, time to do homework after school and frequency of moderate intensity exercise per week associated with poor eyesight of primary and secondary school students. Logistic regression results showed that girls, the distance between the eyes and the computer display screen < 66 cm ($OR = 1.24, 95\% CI = 1.00 \sim 1.52$), and the distance between the eyes and the textbook when reading was less than one foot ($OR = 2.05, 95\% CI = 1.23 \sim 3.40$) were positively associated with poor vision. Low grade students and homework time < 2 hours after school ($OR = 0.76, 95\% CI = 0.62 \sim 0.92$) were negatively associated with lower risk of poor vision. **Conclusion** Educational stage, gender and time of homework after school are associated with the visual acuity of primary and secondary school students in Ningxia. It is suggested that prevention of myopia should start from the low-grade students, reducing after-school work for primary and secondary school students, and increasing outdoor activities. All the girls should be encouraged to engage in more outdoor activities, to prevent the occurrence and development of myopia.

【Key words】 Decision trees; Regression analysis; Vision, low; Prevalence; Students

视力不良指用标准对数视力表检查裸眼远视力时, 单眼或双眼裸眼远视力达不到正常标准, 即低于 5.0^[1]。视力不良作为眼科发病率最高的疾病, 是全世界

界关注的重要公共卫生问题^[2], 世界卫生组织也将列入视觉 2020 行动的目标之一^[3]。中小學生正处于生长发育的关键时期, 不良的用眼习惯及眼负荷过重不仅会导致视力不良, 还会对大脑视皮层功能发育、身体运动能力、情感体验、社会活动等产生负面影响^[4]。2018 年 8 月 30 日, 教育部、国家卫生健康委员会等 8 部门联合印发《综合防控儿童青少年近视实施

【作者简介】 谢小莲 (1994-), 女, 宁夏固原人, 在读硕士, 主要研究方向为儿童青少年生长发育及慢性病防治。

【通讯作者】 赵海萍, E-mail: zhpj@163.com。

DOI: 10.16835/j.cnki.1000-9817.2019.04.025

方案》^[5],首次将近视率作为政府考核的一个标准,目标是到 2030 年将 6 岁儿童近视率控制在 3%,小学生控制在 38% 以下,中学生控制在 60% 以下,高中生控制在 70% 以下。本研究旨在用决策树和 Logistic 回归模型 2 种方法研究宁夏中小学生近视的影响因素,以期更好地控制近视蔓延。

1 对象与方法

1.1 对象 采用分层整群抽样方法,在宁夏银川和吴忠随机抽取 3 所小学、3 所初中、2 所高中,小学按二至五年级分层,每个年级抽取 2 个班;初中按七至八年级分层,每个年级抽取 5 个班;高中按十至十一年级分层,每个年级抽取 5 个班级(一年级因识字率低及六、九、十二 3 个年级因升学压力应答率较低,因此未纳入此次调查),最终确定 3 430 名学生作为研究对象,对学生视力检查和问卷调查。共获得有效问卷 3 257 份,有效率为 94.96%。其中男生 1 581 名(48.5%),女生 1 676 名(51.5%);小学生 794 名(24.4%),中学生 1 397 名(42.9%),高中生 1 066 名(32.7%),年龄 7~18 岁。本研究经过宁夏医科大学伦理委员会的批准。

1.2 方法 问卷内容包括一般人口学资料(性别、家庭所在地等)、视力相关保健行为和运动行为等。视力保健行为分为从不、偶尔、经常、总是 4 个选项,分别表示目标事件发生的频率为 0,40%,60%,100%。

视力检查采用“标准对数视力表灯箱”,视力表的照度 300~500 lx,悬挂高度使视力表与大多数受检对象的双眼呈水平位置。检测结果中双眼裸眼视力≥5.0 判定为正常,单眼或双眼裸眼远视力<5.0 判定为视力不良^[6]。

1.3 统计分析 数据采用 EpiData 3.02 建库录入,所有数据采用双录入并进行逻辑查错,数据统计描述分析采用 SPSS 21.0,百分率比较采用 χ^2 检验,非正态资料分布比较采用秩和检验;影响因素分析采用 Clementine 12.0 中决策树 C 5.0 算法^[7]和 Logistic 回归,检验水准 $\alpha=0.05$ 。

2 结果

2.1 宁夏中小学生视力状况单因素分析 宁夏中小学生视力不良总体检出率为 62.6%,其中男生为 56.8%,女生为 68.1%,差异有统计学意义($P<0.01$);城市和农村中小学生视力不良检出率分别为 63.1%和 62.4%,差异无统计学意义($P=0.68$);小学生视力不良检出率为 41.1%,中学生为 60.0%,高中生为 82.2%,不同学段的学生视力不良检出率差异有统计学意义($P<0.01$),且随着年龄的增长,视力不良检出率递增。此外,每周进行中等强度锻炼天数在视力正常组和视

力不良组之间差异有统计学意义(M 值分别为 3,3, $Z=-2.18,P<0.05$);在近视相关保健行为方面,每天户外活动时间、每日睡眠时间、每日接受光照时间、眼睛距计算机显示屏的距离、读写时胸口离桌边的距离、学习时眼睛距书本的距离、学校下午是否有体育活动和放学后做作业时间等方面差异均有统计学意义(P 值均<0.01)。见表 1。

表 1 宁夏中小学生视力单因素分析

组别	人数	视力不良	χ^2 值	P 值
性别	男	1 581 898(56.8)	44.69	<0.01
	女	1 676 1 142(68.1)		
家庭所在地	城市	1 764 1 113(63.1)	0.18	0.68
	农村	1 299 887(59.4)		
学段	小学生	794 326(41.1)	336.07	<0.01
	初中生	1 397 838(60.0)		
	高中生	1 066 876(82.2)		
每天户外活动时间/h	<1	1 602 1 058(66.0)	20.18	<0.01
	≥1	1 655 982(59.3)		
每日睡眠时间/h	<9	2 880 1 866(64.8)	49.48	<0.01
	≥9	377 174(46.2)		
每日接受光照时间/h	<1	1 906 1 224(64.4)	28.02	<0.01
	≥1	1 351 816(60.4)		
眼睛距计算机的距离/cm	<66	748 522(69.8)	21.22	<0.01
	≥66	2 509 1 518(60.5)		
读写时胸口离桌边不足 1 拳	从不	308 163(52.9)	24.37	<0.01
	偶尔	1 856 1 142(61.5)		
	经常	884 588(66.8)		
	总是	209 147(73.9)		
学习时眼睛距书本不足 1 尺(33 cm)	从不	358 176(49.2)	45.06	<0.01
	偶尔	1 776 1 101(62.0)		
	经常	935 624(66.7)		
	总是	188 139(73.9)		
写字时手指距笔尖 1 寸	从不	704 444(63.1)	0.37	0.95
	偶尔	1 318 823(62.4)		
	经常	836 519(62.1)		
	总是	399 254(63.7)		
学校下午是否有体育活动	否	1 690 1 125(66.6)	23.23	<0.01
	是	1 567 915(58.4)		
放学后做作业时间/(h·d ⁻¹)	<1	181 88(48.6)	64.45	<0.01
	1~<2	772 409(53.0)		
	≥2	2 304 1 543(67.0)		

注:() 内数字为检出率/%。

2.2 宁夏中小学生视力不良 Logistic 回归分析 以是否视力不良(1=是,2=否)作为因变量,将人口学资料、视力相关保健行为、运动行为 3 个方面单因素分析差异具有统计学意义的 11 个因素作为自变量进行二分类 Logistic 回归分析,结果显示,性别(0=男生,1=女生)、学段(1=高中,2=初中,3=小学)、眼睛距计算机显示屏的距离(1=≥66 cm,2=<66 cm)、学习时眼睛距书本的距离(1=从不,2=偶尔,3=经常,4=总是)和每日放学后做作业时间(1=≥2 h,2=1~<2 h,3=<1 h)与中小学生视力不良有关(P 值均<0.05)。与参照水平相比,性别女、眼睛距计算机显示屏距离<66 cm、学习时眼睛距课本不足 1 尺(33 cm)与视力不良发生呈正相关;低学段组和每日放学后做作业时间<2 h 与视力不良发生呈负相关。见表 2。

表 2 宁夏中小學生视力不良影响因素 Logistic 分析 (n=3 257)

影响因素		β 值	标准误	Wald χ^2 值	P 值	OR 值 (OR 值 95%CI)
性别	女	0.50	0.09	34.11	<0.01	1.65 (1.40~1.96)
学段	初中	-1.04	0.11	88.57	<0.01	0.35 (0.28~0.44)
	小学	-1.83	0.13	188.64	<0.01	0.16 (0.12~0.21)
眼睛距离计算机显示屏距离/cm	<66	0.21	0.11	3.92	<0.05	1.24 (1.00~1.52)
学习时眼睛距书本不足 1 尺 (33 cm)	偶尔	0.35	0.15	5.79	0.02	1.42 (1.07~1.89)
	经常	0.43	0.16	7.26	0.01	1.55 (1.13~2.13)
	总是	0.72	0.26	7.66	0.01	2.05 (1.23~3.40)
放学后做作业时间	1~<2	-0.28	0.10	7.70	0.01	0.76 (0.62~0.92)
/(h·d ⁻¹)	<1	-0.52	0.18	8.46	0.00	0.59 (0.42~0.84)

2.3 宁夏中小學生视力不良的决策树分析 将上述单因素分析结果中有意义的指标纳入决策树模型,模型共 12 个节点,7 个叶节点。由图 1 可知,对视力不良影响最大的因素是学段分组,学段越高,近视不良检出率越高,小学、初中、高中分别占该节点的 42.6%, 59.4%, 83.2%;性别因素是小学生和初中生视力不良

的一个重要影响因素,女生的视力不良检出率均高于男生,分别占该节点的 50.6%和 65.8%;在中学男生中,每日做作业时间 ≥ 2 h,视力不良检出率最高,占该节点的 57.7%;在小学女生中,每周进行中等强度锻炼的频率较低者,视力不良检出率较高,占该节点的 56.9%。

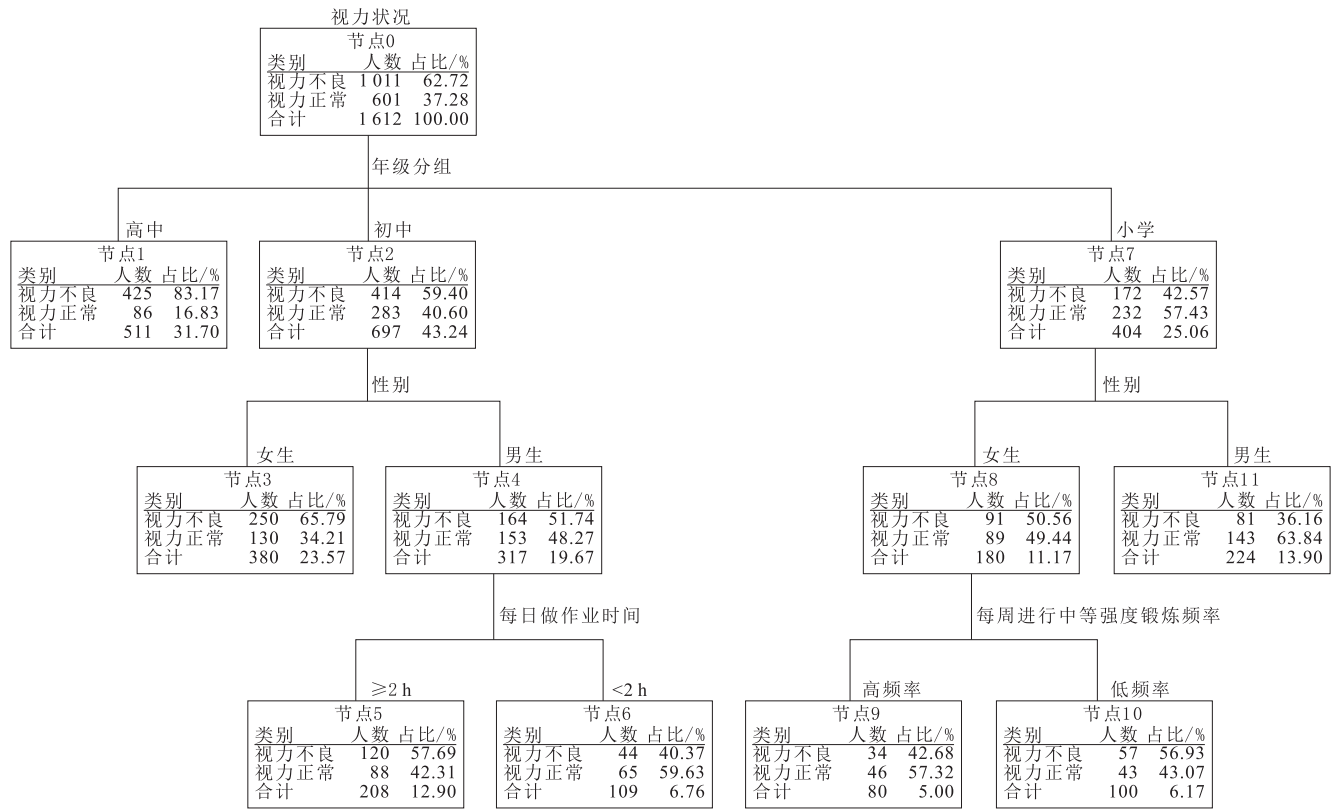


图 1 宁夏中小學生视力不良决策树

2.4 Logistic 回归模型与决策树模型的评价 2 个模型预测结果显示,Logistic 回归模型灵敏度 (71.1%) 高于决策树 C 5.0 模型 (70.8% 和 69.0%), 特异度 (60.6%) 高于决策树测试集 (58.5%) 且低于决策树训练集 (61.6%), 2 个模型正确分类百分比均 $\geq 70\%$ 。说明两者评价效能相近,Logistic 灵敏度高,决策树 C 5.0 特异度高,两者结合更有利于分析视力不良的影响因素。见表 3。

表 3 中小學生视力不良 Logistic 回归模型和决策树模型判断结果与评价

模型	预测模型 判断结果	视力标准判断结果		灵敏度 /%	特异度 /%	正确分类 百分比/%
		视力不良	视力正常			
决策树模型	训练集	视力不良	857	154	70.8	70.0
		视力正常	354	247		
	测试集	视力不良	870	159	69.0	70.6
		视力正常	392	224		
Logistic 回归	视力不良	1 434	282	71.1	60.6	70.6
	视力正常	582	433			

3 讨论

本次研究结果显示,宁夏中小學生视力不良总体检出率为 62.6%,远高于 2010 年的 39.6%^[8];男生视力不良检出率为 56.8%,女生为 68.1%,女生视力不良检出率高于男生,与其他学者的研究一致^[8-10]。可能由于女生更喜静,且学习更刻苦,使户外活动时间相对减少,从而视近时间延长,使睫状肌处于一个长期的紧张状态,易导致睫状肌紧张、痉挛,晶状体变凸,眼轴变长,形成视力不良^[10]。小学生视力不良检出率为 41.1%,中学生为 60.0%,高中生为 82.2%,与 2010 年^[11]全国学生体质健康调研数据(小学生 48.8%,中学生 75.9%,高中生 79.2%)相比,宁夏小学生和中学生视力不良检出率稍低于全国水平,而高中生高于全国水平,提示宁夏中小學生视力状况十分严峻;不同学段的学生视力不良检出率差异有统计学意义,且随着年龄的增长递增,与程广印等^[12-14]的研究一致,随着年级的升高,学生学习任务逐步加重,导致用眼时间延长,而且随着年龄增长,户外活动也随之减少,导致近视率增多^[13]。也有学者研究表明,随着年龄的增长,高度视力不良的比例逐渐增高^[9,14],提示预防近视的发生发展必须从低年龄组开始。

决策树 C 5.0 算法和 Logistic 回归分析结果均显示,学段、性别、放学后做作业时间是中小學生视力不良最主要的影响因素。作业时间越长患视力不良的可能性越大,与视近时间过长有关。此外,Logistic 回归分析结果显示,与参照水平相比,眼睛距计算机显示屏距离<66 cm、读书时眼睛距离课本不足 1 尺与视力不良呈正相关,与丁婷婷等^[15]的研究相似。原因也与长时间近距离阅读有关,研究表明,长时间近距离阅读可使睫状肌持续收缩,如果中间不通过休息来放松睫状肌,便会导致睫状肌长时间的调节痉挛,使巩膜逐渐延伸、眼轴变长,产生近视或近视程度加深^[12]。此外,决策树结果还显示,每周进行中等强度锻炼的频率也是视力不良的影响因素,运动不仅能使眼外肌与眼内调节肌群的调节能力增强,还能使眼球血液循环增强,血供增加的同时,眼球的抗张能力增强,从而阻止了眼球轴的增长,一方面有效缓解了眼部疲劳,另一方面起到了锻炼眼球的目的^[16]。

决策树和 Logistic 模型在预测近视影响因素方面的正确分类百分比均高于或等于 70%,模型效果较好^[17]。Logistic 回归模型显示了近视相关因素与视力不良的数量依存关系,在了解自变量对因变量的贡献方面优于决策树模型,而决策树模型可以深入数据内部,分析数据在不同水平下的预测概率,如在学段分组结果中进一步分析了不同性别初中生发生视力不良的概率,预测出该年级女生发生视力不良的概率较高,提示应着重关心初中生女生的视力状况^[18]。本次研究中,决策树和 Logistic 回归模型分析得出的影响

因素不完全相同,建议在影响因素的分析中将 2 个模型结合使用,更有利于挖掘出潜在的危险因素。

综上所述,宁夏中小學生视力不良现状堪忧,急需有关部门做出相关预防措施,并将工作重点放在低学段组,尽可能减少中小學生视近和课后作业时间,适量增加户外活动时间,同时建议老师多鼓励女生进行户外活动,学校也应推出适合女生的户外体育项目,从而预防近视的发生和发展。此外,决策树 C 5.0 模型和 Logistic 回归模型在近视影响因素分析中的预测效果相当,两者预测结果可以互相补充,结合使用可以更充分地分析视力不良的影响因素。

4 参考文献

- [1] 孙凯丽.中小學生近距离工作现状与视力不良关系及预防对策的研究[D].天津:天津医科大学,2017.
- [2] 木尼热·热孜.伊宁市维汉中小學生近视相关环境及遗传因素的调查[D].乌鲁木齐:新疆医科大学,2017.
- [3] ACKLAND P,石磊.关于“视觉 2020”防盲计划的研究[J].实用防盲技术,2011,6(4):139-140.
- [4] 程绪婷,王宏,窦义蓉.重庆主城区视力不良小學生视力相关生存质量与视力保健行为关系[J].中国学校卫生,2016,37(9):1311-1313.
- [5] 樊泽明,刘立京,王海涛.扎实推进全国儿童青少年近视防控工作[J].中国学校卫生,2018,39(11):1605-1611.
- [6] 佚名.学校标准对数视力表使用标准[EB/OL].[2018-06-30].<https://wenku.baidu.com/view/d9296a5cf46527d3240ce0ea.html?from=search>.2019.
- [7] 熊蜀峰,聂黎明.基于 C5.0 算法的学成绩分析决策树构造[J].科技信息,2010(8):24-25.
- [8] 许红霞,石丽文,牛晓丽,等.宁夏地区 2005—2014 年中小學生视力不良发展趋势[J].中国学校卫生,2016,37(3):412-413.
- [9] 余家麟,曾金水,邱毅,等.2017 年深圳市宝安区小學生视力不良流行现状及影响因素分析[J].实用预防医学,2018,25(6):740-743.
- [10] CHARMAN W N.Myopia, posture and the visual environment[J].Ophthalmic Physiol Opt,2011,31(5):494-501.
- [11] 刘玮.兰州市 2010 与 2014 年中小學生体制与健康状况对比研究[D].兰州:兰州大学,2018.
- [12] 马翠侠,付玲玲,汪丽娟,等.合肥市小學生近视现状及相关危险因素的分析[J].临床眼科杂志,2018,26(3):265-268.
- [13] 张湘雯,屈艳梅,张兰英.北京市海淀区小學生近视现状调查与影响因素分析[J].国际眼科杂志,2018,18(8):1477-1479.
- [14] 程广印,李步云.北京市西城区 2011—2012 年中小學生视力不良状况[J].中国学校卫生,2014,35(5):781-782.
- [15] 丁婷婷,雷佳莞,黄华平.苍南县 1612 名高考学生近视状况及影响因素分析[J].中国农村卫生事业管理,2018,38(5):596-598.
- [16] 陈远莉,刘琳琳,丁南杰.乒乓球运动处方对改善小學生视力的运用与实践研究[J].体育世界,2018,11(11):194-195.
- [17] 夏莉莉,赵华硕,徐渭,等.决策树和回归技术在超重危险因素分析中的应用[J].现代预防医学,2018,45(9):1549-1552.
- [18] 刘兵,李苹,朱玫桦,等.决策树模型与 Logistic 回归模型在胃癌高危人群干预效果影响因素分析中的应用[J].中国卫生统计,2018,35(2):70-73.

收稿日期:2019-01-21;修回日期:2019-02-27