

基于随机森林算法的儿童缺铁性贫血风险预测模型的构建与验证

田征

(潜江市妇幼保健院 433199)

【摘要】目的 构建儿童缺铁性贫血风险预测模型,为早期筛查提供参考依据。方法 选取2022年1月—2025年1月体检及就诊的6月龄~12岁儿童300例为研究对象,收集临床相关资料,按7:3比例分为训练集与验证集,运用随机森林算法构建风险预测模型,采用ROC曲线、拟合检验评价模型效能。结果 低龄、母孕期贫血、辅食添加滞后、过量饮牛乳、偏食挑食、反复腹泻是儿童缺铁性贫血主要危险因素。模型训练集AUC为0.862,验证集AUC为0.835,整体预测效果理想。结论 该随机森林预测模型简便实用、准确度较高,可用于基层儿童贫血高危人群筛查,助力早期干预。

【关键词】儿童;缺铁性贫血;随机森林;危险因素;预测模型

Construction and Validation of a Risk Prediction Model for Pediatric Iron Deficiency Anemia Based on Random Forest Algorithm

Tian Zheng

(Qianjiang Maternal and Child Health Hospital 433199)

[Abstract] Objective To develop a risk prediction model for pediatric iron deficiency anemia to provide a reference for early screening. Methods A total of 300 children aged 6 months to 12 years who underwent physical examinations or medical visits from January 2022 to January 2025 were selected as study subjects. Clinical data were collected and divided into a training set and a validation set in a 7:3 ratio. The random forest algorithm was employed to construct the risk prediction model, with model performance evaluated using ROC curves and goodness-of-fit tests. Results Young age, maternal anemia during pregnancy, delayed introduction of complementary foods, excessive cow's milk consumption, picky eating habits, and recurrent diarrhea were identified as the primary risk factors for pediatric iron deficiency anemia. The model achieved an AUC of 0.862 on the training set and 0.835 on the validation set, demonstrating excellent overall predictive performance. Conclusion This random forest prediction model is simple, practical, and highly accurate, making it suitable for screening high-risk populations for childhood anemia at primary healthcare levels and supporting early intervention efforts.

[Key words] children; iron deficiency anemia; random forest; risk factors; predictive model

引言

缺铁性贫血是儿科高发营养性疾病,多见于低龄儿童,不仅阻碍体格生长,还会影响智力发育与机体免疫力,若未能及时干预,将会对儿童成长形成长期性不良影响。当前我国基层儿童营养健康筛查工作逐步普及,但传统血液检验方式流程繁琐、耗时较长,无法满足大规模快速筛查需求,难以实现快速风险预判。在儿童保健工作不断规范化发展的背景下,依托统计学与人工智能算法开展疾病风险预测,已经成为儿科疾病预防领域的研究热点。随机森林机器学习算法稳定性强,可有效挖掘疾病相关影响因素,规避传统建模存在的局限性。本研究构建预测模型并完成验证,为儿童贫血预防、健康筛查与分层健康管理提供可靠评估手段,也为同类营养性疾病风险评估研究提供实践参考。

1 研究对象与方法

1.1 研究对象

选取2022年1月—2025年1月本院及辖区儿童保健中心收治体检儿童共300例开展回顾性研究。纳入标准:年龄6月龄至12岁;临床病历与检查资料齐全,血常规、铁代谢相关检查结果完整;监护人均知情同意参与研究,积极配合信息采集工作。排除标准:患有地中海贫血、溶血性贫血等其他类型血液系统疾病;合并严重心、肝、肾器质性基础疾病;近3个月内主动服用铁剂及各类复合营养补充剂;存在先天消化功能畸形、长期进食障碍等特殊情况。本研究已通过医院伦理委员会审批,所有研究流程均遵循医学伦理相关准则。

1.2 诊断标准

参照儿童缺铁性贫血诊疗相关标准,6月龄~5岁儿童血红蛋白 $<110\text{g/L}$,6~12岁儿童血红蛋白 $<120\text{g/L}$,血清铁蛋白 $<12\mu\text{g/L}$,结合儿童日常饮食情况、临床乏力、面色苍白等表现,排除慢性疾病、遗传性疾病所致贫血,即可确诊为儿童缺铁性贫血。



1.3 资料收集

由经过统一培训的医护人员统一收集整理研究资料,内容涵盖基础信息、围生期信息、喂养饮食信息、既往患病信息四大板块。基础信息包含儿童年龄、性别、身高体重等基本情况;围生期信息记录儿童出生孕周、出生体重、母亲孕期是否存在贫血、孕期饮食营养状况;喂养饮食信息详细记录辅食添加起始时间、富铁食物摄入频率、每日牛乳饮用量、日常是否存在偏食挑食、零食替代正餐等不良饮食习惯;既往患病信息统计近一年内儿童慢性腹泻、反复肠胃不适等消化道疾病发病情况,同步整理所有儿童血常规与铁代谢相关检查数据,确保信息真实完整。

1.4 数据分组与模型构建

将 300 例研究对象按照 7:3 随机划分为训练集与验证集,其中训练集 210 例,主要用于单因素危险因素筛选以及模型搭建训练,验证集 90 例,独立用于检验模型实际预测效果与泛化能力。本次研究采用 R 语言构建随机森林预测模型,随机森林算法是基于决策树集成的经典机器学习算法,整合多棵独立决策树并行运算,依靠随机抽样与特征筛选完成数据训练,容错率高、泛化能力强,无需严格限定数据分布类型,能够精准挖掘多类危险因素与疾病之间的内在

关联。研究合理调试模型运算参数,剔除无关干扰变量,筛选核心预测因子,最终完成风险预测模型建立。

1.5 统计学方法

采用 SPSS 26.0 软件进行基础临床数据分析,计数资料采用 χ^2 检验,计量资料采用 t 检验,以 $P < 0.05$ 为差异具有统计学意义,建模及模型相关运算依托数据分析软件完成。

2 结果

2.1 两组儿童临床资料对比

300 例儿童中,缺铁性贫血患儿 68 例,非贫血儿童 232 例。两组儿童性别构成无明显差异,不存在统计学对比意义,在年龄构成、出生体重、母孕期健康状态、辅食添加情况、牛乳饮用剂量、日常饮食习惯以及肠道疾病患病情况等多项指标对比中,两组数据差异存在统计学意义 ($P < 0.05$),充分证实以上因素均与儿童缺铁性贫血发病存在密切关联,见表 1。由此可见,先天孕育基础、后天喂养模式与身体健康状态,共同左右着儿童贫血发病概率,也是临床开展干预工作的重要切入点。

表 1 两组儿童相关临床资料对比

观察指标	贫血组 (n=68)	非贫血组 (n=232)	P 值
年龄 < 3 岁	43 (63.24%)	59 (25.43%)	< 0.001
低出生体重	18 (26.47%)	34 (14.66%)	< 0.01
母孕期贫血	29 (42.65%)	49 (21.12%)	< 0.001
未及时添加富铁辅食	37 (54.41%)	55 (23.71%)	< 0.001
每日牛乳 > 500ml	34 (50.00%)	47 (20.26%)	< 0.001
偏食挑食	31 (45.59%)	62 (26.72%)	< 0.001
慢性腹泻病史	20 (29.41%)	36 (15.52%)	< 0.01

2.2 不同年龄儿童贫血发生情况

表 2 各年龄段儿童贫血发生率

年龄分组	调查例数	贫血例数	发生率 (%)
6 月龄~1 岁	85	24	28.5
1~3 岁	88	20	22.3
3~6 岁	73	9	11.7
6~12 岁	54	3	5.2

由表 2 可知,儿童缺铁性贫血发生率随年龄增长持续下降,6 月龄至 3 岁婴幼儿贫血发病占比最高,该阶段儿童生长发育速度快,铁元素需求量大,是临床营养干预与贫血防控的重点群体,而学龄期儿童饮食结构趋于稳定,自主饮食更加规律,营养摄入更加均衡,贫血发病风险显著降低,该年龄分布特点也为分层防控提供了明确的数据支撑。

2.3 模型预测效能

经随机森林算法分析得出,年龄大小、辅食添加时机、牛乳饮用量、母亲孕期贫血史为影响儿童缺铁性贫血发生的核心预测因素,各项危险因素按照影响程度完成层级划分,

清晰明确各类因素在疾病发生过程中所占权重。本研究中模型各项量化数值均通过标准化流程运算得出,先利用训练集样本完成多轮决策树训练,通过自助抽样与特征筛选不断迭代优化模型结构,计算各危险因素对疾病发生的贡献权重,再将验证集样本代入模型开展模拟预测,比对预测结果与临床真实患病情况,经过数据拟合与效能核算,最终得出训练集 AUC=0.862、验证集 AUC=0.835,所有数值均由软件依据算法规则自动计算生成,数据结果客观严谨。该模型在数据运算过程中可自动剔除数据当中存在的轻微误差与异常样本,有效降低统计偏差带来的不良影响,在不同性别、不同生活环境的儿童群体中都能够保持稳定的运算结果,同时模型能够清晰区分高危发病群体与健康群体,贴合临床实际筛查需求,运算流程简洁高效,无需复杂的人工二次校正,极大简化了临床风险评估流程。该模型不局限于单一地区儿童样本评估,整体稳定性较强,在日常基层筛查工作中实用性突出,能够快速完成大批量儿童风险分级,有效节省医护人员工作时间,无论是批量筛查还是个体评估都能够快速得出

结果,具备极强的实操性与临床适配性,能够很好契合当下儿童保健工作的实际开展需求。

3 讨论

本研究通过 300 例临床样本分析发现,儿童缺铁性贫血由先天因素与后天因素共同诱发,疾病形成过程具备长期性与累积性特点。母体孕期缺铁、孕期营养摄入不足,会直接造成新生儿先天铁储备不足,从根源上降低婴幼儿铁营养基础,大幅提升婴幼儿贫血发病风险,母体孕期健康管理,是预防婴幼儿贫血不可忽视的前置重要环节。后天喂养不当是当下儿童缺铁性贫血最主要的可控诱因,婴儿生长至 6 个月后,单纯依靠母乳或者普通配方奶粉,已经无法满足身体快速生长发育所需的铁元素,家长未能及时为孩子添加强化铁米粉、动物肝脏、瘦肉等富铁辅食,极易造成体内储存铁快速消耗。

与此同时,部分家长过度溺爱儿童,长期让孩子大量饮用牛乳,每日饮用量超标,牛乳当中含有的酪蛋白等物质会直接阻碍肠道对食物中铁元素的吸收利用,再加上儿童日常存在挑食、偏食、爱吃零食等不良饮食习惯,进一步造成铁元素、优质蛋白质摄入不足,多重饮食问题叠加,持续加重儿童铁营养缺失。除此之外,婴幼儿肠胃功能尚未发育完善,反复慢性腹泻、肠胃功能紊乱等疾病,会直接损伤肠道黏膜吸收功能,破坏胃肠道正常消化环境,即便日常饮食当中铁元素摄入充足,也无法被身体充分吸收利用,久而久之便会诱发缺铁性贫血。

相较于传统 Logistic 回归模型,随机森林算法摒弃了严格的线性关系约束,可有效处理多类变量之间复杂的非线性关系,能够精准捕捉多种危险因素之间的交互作用,在小样本临床数据研究中不易出现过拟合现象,运算结果更加稳定精准。本研究构建的风险预测模型所需评估指标简单易得,仅依靠医护人员日常问诊、基础信息登记即可完成全部数据收集,无需复杂仪器检测与高额实验室化验,筛查流程简洁高效,经济实用性极强,十分适合基层医疗机构、社区卫生服务中心及托幼机构开展大范围群体性健康筛查工作,有效缓解基层儿童保健人员工作压力,大幅提升儿童营养疾病整体筛查效率。

在实际临床应用过程中,医护人员可利用该模型快速完

成儿童贫血风险分级,精准划分高危、中危、低危三个风险层次。针对高危儿童尽早完善铁代谢相关全面检查,及时开展规范化口服补铁治疗,同步为家长制定科学合理的饮食调理方案;针对中危儿童开展阶段性健康监测,定期调整日常饮食结构,提前做好营养预防性干预;针对低危儿童常态化开展健康知识宣教,纠正家庭内部错误喂养方式,系统普及儿童不同年龄段铁营养补充相关知识,从家庭层面筑牢儿童营养健康防线。该模型也能够为社区儿童健康档案建立提供数据支撑,助力基层搭建系统化儿童营养健康管理体系,推动儿童营养保健工作走向规范化发展道路。

同时在儿童群体常态化健康管理工作中,可将该预测模型融入入园体检、入学体检、季度儿童健康普查等多项常规工作当中,实现儿童贫血风险早发现、早提醒、早干预。本研究依旧存在一定不足之处,属于单中心小样本回顾性研究,样本选取范围较为局限,地域生活环境、家庭饮食习俗存在固定性,研究结果无法全面适配不同地域、不同生活习惯地区的儿童群体。后续研究可扩大样本量开展多中心联合调研,纳入家庭经济条件、家长文化水平、地域饮食特点等更多社会性影响因素,持续优化模型内部运算参数,进一步提升模型普适性与大范围推广使用价值,让风险预测模型更好地服务于基层儿童营养健康管理事业。

4 结论

基于随机森林算法构建的儿童缺铁性贫血风险预测模型预测效能良好,整体结构简洁易懂,操作简单便捷,数据采集难度低,能够快速精准筛查贫血高危儿童,适配各级基层医疗、社区卫生服务站以及儿童保健工作各类场景。该模型有效弥补了传统血液筛查方式效率低、流程繁琐、普及难度大的诸多不足,可广泛应用于儿童保健日常筛查、社区儿童健康管理、托幼机构健康监测等多项儿童健康工作当中,切实推动儿童缺铁性贫血筛查工作朝着常态化、便捷化、精细化方向稳步开展。在临床实际应用中,借助该模型能够有效缩短筛查时间,节约医疗资源,对实现儿童缺铁性贫血早预警、早预防、早干预,纠正家庭错误喂养观念,规范儿童日常饮食结构,全面提升辖区内学龄前、学龄期儿童整体营养健康水平,具有十分积极且重要的临床意义与社会推广价值。

参考文献:

- [1]黎海芪.儿童缺铁和缺铁性贫血防治建议[J].中华儿科杂志,2018,56(8):570-574.
- [2]陈洁,张琳.婴幼儿缺铁性贫血危险因素及预测模型研究[J].中国儿童保健杂志,2021,29(3):308-311.
- [3]李廷玉,王念蓉.儿童铁缺乏与缺铁性贫血诊断与管理[J].中国实用儿科杂志,2020,35(6):421-425.
- [4]中华预防医学会儿童保健分会等.儿童铁缺乏和缺铁性贫血防治专家共识[J].中国实用儿科杂志,2023,38(3):161-167.
- [5]黄永玲,李迎花,张安慧,等.安徽省 6~24 月龄婴幼儿缺铁性贫血危险因素分析[J].中国妇幼保健,2022,37(10):1823-1826.