

DOI: 10.19296/j.cnki.1008-2409.2024-05-009

• 论 著 •

• ORIGINAL ARTICLE •

## XGBoost 模型与 Logistic 回归模型预测儿童 TBTB 合并 MPP 的价值

刘小峰 洪智文 陈芳 严灿 谢齐放

( 南华大学衡阳医学院附属长沙中心医院儿科, 长沙 410004 )

**摘要** 目的 比较 XGBoost 模型和 Logistic 回归模型在预测儿童气管支气管结核( TBTB ) 合并肺炎支原体肺炎( MPP ) 中的效能。方法 收集 2018 年 6 月至 2023 年 1 月入院治疗的 212 例 MPP 患儿临床资料 根据是否合并 TBTB 将其分为合并组(  $n=42$  ) 与肺炎组(  $n=170$  ) 。通过 XGBoost 和 Logistic 回归构建儿童 TBTB 合并 MPP 的临床预测模型, 采用受试者工作特征( ROC ) 曲线绘制两个模型在预测儿童 TBTB 合并 MPP 中的预测价值; 采用 Delong 检验确定两个模型曲线下面积的差异。结果 XGBoost 模型筛选符合情况的相关变量为地区、咳嗽时间、发热史、咳嗽史、SIRI、WBC、湿罗音、SII, 多因素 Logistic 回归分析发现 地区、发热史、咳嗽史、咳嗽时间、WBC 是儿童 TBTB 合并 MPP 的独立危险因素(  $P<0.05$  ), XGBoost 曲线下面积为 0.996, 而 Logistic 回归模型曲线下面积仅为 0.851。Delong 检验发现 XGBoost 模型预测效能高于 Logistic 回归模型, 差异有统计学意义(  $P<0.01$  ) 。结论 XGBoost 模型优于 Logistic 回归模型, 可作为诊断儿童 TBTB 合并 MPP 的有效工具。

**关键词:** 气管支气管结核; 肺炎支原体肺炎; XGBoost; Logistic 回归; 预测模型

中图分类号: R725.6

文献标志码: A

文章编号: 1008-2409( 2024 ) 05-0055-07

### Value of XGBoost model and Logistic regression model in predicting children with TBTB ambined with MPP

LIU Xiaofeng , HONG Zhiwen , CHEN Fang , YAN Can , XIE Qifang

( Department of Pediatrics , the Affiliated Changsha Central Hospital , Hengyang Medical School , University of South China , Changsha 410004 , China )

**Abstract Objective** To compare the predictive value of XGBoost model and Logistic regression model in children with tracheobronchial tuberculosis ( TBTB ) with mycoplasma pneumoniae pneumonia( MPP ) . **Methods** The clinical data of 212 children with MPP , admitted from June 2018 and January 2023 , were collected. Based on whether the children were diagnosed with TBTB , they were divided into a combind

基金项目: 湖南省卫生健康委科研计划项目( D202306016238 ) 。

第一作者: 刘小峰, 硕士研究生, 主治医师, 研究方向为儿童呼吸系统疾病; 洪智文, 硕士研究生, 主治医师, 研究方向为小儿呼吸系统疾病。

通信作者: 严灿 215939094@qq.com; 谢齐放 fangon@163.com。

group( $n=42$ ) and a pneumonia group( $n=170$ ). clinical prediction model for pediatric TBTB combined with MPP was constructed using XGBoost and Logistic regression. The receiver operating characteristic (ROC) curves was used to plot the predictive value of two models in predicting the of pediatric TBTB combined with MPP. The Delong test was used to determine the difference in the area under the curve between the two models. **Results** The relevant variables that matched the situation and filtered with the XGBoost model were region , cough duration , history of fever , history of cough , SIRS , white blood cell count ( WBC ) , moist rale , and SII. Multivariate Logistic regression analysis revealed that region , fever history , cough history , cough duration , and WBC were risk factors for TBTB complicated with MPP in children ( $P<0.05$ ). The area under the curve of the XGBoost model was 0.996 , while the area under the curve of the Logistic regression model was only 0.851. Delong's test found that the predictive efficacy of the XGBoost model was significantly higher than that of the Logistic model , with a statistically significant difference( $P<0.01$ ). **Conclusion** The XGBoost model is better than the Logistic regression model , and can be used as an effective tool for the diagnosis of TBTB combined with MPP in children.

**Keywords:** tracheobronchial tuberculosis; mycoplasma pneumoniae pneumonia; XGBoost; Logistic regression; predictive modeling

儿童气管支气管结核(tracheobronchial tuberculosis , TBTB)和肺炎支原体肺炎(mycoplasma pneumoniae pneumonia , MPP)是临床中较为常见的两种儿童呼吸道疾病<sup>[1]</sup>。TBTB是一种特殊类型的肺结核,主要累及气管和支气管的黏膜及周围组织,可能导致气道狭窄、闭塞和肺实变等严重并发症,从而显著影响患儿的呼吸功能<sup>[2-3]</sup>。MPP是我国5岁及以上儿童最主要的社区获得性肺炎,其临床表现和影像学特征常呈现肺间质性、小叶性、大叶性肺炎和肺门淋巴结肿大等多样性<sup>[4]</sup>。然而,当TBTB和MPP同时存在时,由于两者在症状表现和影像学特征上的重叠,如发热、咳嗽、呼吸困难及肺实变或肺浸润等,初步鉴别诊断往往存在较大难度,易造成误诊或漏诊<sup>[5]</sup>。这不仅延误患儿的治疗时机,还可能导致疾病的进一步恶化,严重影响患儿的预后和生存质量。因此,如何准确区分TBTB合并MPP与单一疾病的发生,对于临床医生制定合理的治疗策略至关重要<sup>[6]</sup>。随着人工智能和机器学习技术在医学领域的不断应用,利用数据驱动的预测模型来辅助临床诊断已成为一种趋势<sup>[7]</sup>。XGBoost模型作为一种先进的机器学习算法,因其强大的特征选择能力和出色的预测性能,逐渐被应用于多种疾病的诊断和预后评估中<sup>[8]</sup>。而Logistic回归模型作为传统的统计学方法,依然在医学预测模型中占据重要地位。本研究通过

比较XGBoost模型与Logistic回归模型在预测儿童TBTB合并MPP中的效能,探讨两种模型在诊断中的应用价值,以期为临床医生提供更为精准的诊断工具,进而改善患儿的治疗效果和预后。

## 1 方法与资料

### 1.1 一般资料

选取南华大学衡阳医学院附属长沙中心医院2018年6月1日至2023年1月1日收治的212例MPP患儿临床资料,根据是否合并TBTB将其分为合并组( $n=42$ )与肺炎组( $n=170$ )。本研究通过中国临床试验注册中心的临床注册(临床批号:ChiCTR2300076648)。本研究通过湖南省长沙市中心医院医学伦理委员会批准(伦理批号:2023-057)。

纳入标准:①临床诊断为MPP的儿童<sup>[4]</sup>;②均接受常规检查;③临床病历资料完整;④在本研究之前未接受其他治疗。

排除标准:①伴有严重心、肝、肾功能障者;②伴有其他肺部疾病;③伴有自身免疫病;④伴有循环系统障碍;⑤伴有恶性肿瘤;⑥合并其他病原微生物感染;⑦拒绝参与或不能配合研究。

### 1.2 诊断标准

气管支气管结核的诊断标准参考中华人民共和

国卫生行业标准——肺结核诊断(WS288-2017)<sup>[9]</sup>。根据该标准,通过进行纤维支气管镜检查可以直接观察到气管和支气管的病变情况。如果活检病理、分泌物的结核分枝杆菌涂片、培养、核酸检测中的任一结果呈阳性,即可诊断为气管支气管结核。

MPP 的诊断标准参考《诸福棠实用儿科学》<sup>[10]</sup>。该书指出,患儿的临床表现包括咳嗽、发热、呼吸困难等症状。体检时可以听到局部呼吸音减弱、干湿啰音,病灶部位可能会出现实音或浊音。肺部 CT 检查显示肺实质炎症浸润病变或节段性大片状密度增浓影。

### 1.3 临床资料收集

通过患儿的电子病历和门诊复查记录获取患儿的临床资料。临床资料包括年龄、性别、地区、发热史、咳嗽史、咳嗽时间、湿啰音、白细胞(WBC)、中性粒细胞(NE)、淋巴细胞(LYM)、血红蛋白(HGB)、血小板(PLT)、单核细胞(MONO)、C 反应蛋白(CRP)、乳酸脱氢酶(LDH)、中性粒细胞与淋巴细胞比值(NLR)、血小板与淋巴细胞比值(PLR)、单核细胞与淋巴细胞比值(MLR)、系统性免疫炎症指数(SII)和系统性免疫炎症指数相关指标(SIRI)。NLR=NEU/LYM; PLR=PLT/LYM; MLR=MONO/LYM; SII=PLT\*(NEU/LYM); SIRI=(NEU\*MONO)/LYM。

### 1.4 模型构建

XGBoost 模型是一种用于二分类分析的算法。本研究采用 XGBoost 模型对名为“data.txt”的数据进行分析。为使数据适用于二分类分析,将数据中的 Subgroups 列转换为二进制标签作为目标变量。XGBoost 模型具体使用“gbtree”作为提升器,使用二分类逻辑回归作为目标函数,并将 AUC 作为评估指标。本研究采用一系列训练模型,其中最大深度为 6,学习率为 0.01,样本率和列样本率均为 0.8。设置的训练终止条件是连续 10 次迭代没有性能提升。在完成模型训练后,对其性能进行评估,绘制受试者工作特征(ROC)曲线并计算 AUC 值。此外,还对数据特征的重要性进行分析,并将结果可视化展示。Logistic 回归模型是一种用于二分类或多分类问题的线性模型。相比于线性回归,Logistic 回归模型估

计的是一个概率,用来描述给定输入特征下某一事件发生的可能性。模型的输出经过 sigmoid 函数或 softmax 函数(用于多分类)映射到 [0, 1] 区间内,然后根据这个概率进行分类。

### 1.5 结果测量

通过 XGBoost 和 Logistic 回归模型构建儿童 TBTB 合并 MPP 的临床预测模型,采用 ROC 曲线绘制两个模型在预测儿童 TBTB 合并 MPP 患儿中的预测价值,采用 Delong 检验确定两个模型 AUC 的差异。

### 1.6 统计学方法

采用 SPSS 26.0 统计软件处理数据,计数资料以百分比(%)表示,用  $\chi^2$  检验;计量数据用 Kolmogorov-Smirnov 对数据分布情况进行统计,符合正态分布的数据以  $(\bar{x} \pm s)$  表示,采用独立样本 *t* 检验,不符合正态分布的计量资料采用中位数 [ $M(P_{25}, P_{75})$ ] 表示,描述数据的集中趋势和离散程度,用秩和检验;采用 Logistic 回归模型分析影响儿童 TBTB 合并 MPP 的危险因素;利用 R 语言 4.2.2 进行统计分析,采用 XGBoost 回归的建模;rocr 包用于 ROC 曲线的制作。 $P < 0.05$  表示差异具有统计学意义。

## 2 结果

### 2.1 XGBoost 特征资料筛选

对患儿数据采用 XGBoost 模型,得出相关变量与该模型之间的权重关系,再根据权重进行划分,剔除权重小于 1%的变量,最后得到符合情况的相关变量共 8 个特征变量,分别是地区、咳嗽时间、发热史、咳嗽史、SIRI、WBC、湿罗音、SII,如表 1 所示。

表 1 XGBoost 特征资料分布情况

因素	占比/%
地区	54.66
咳嗽时间	25.81
发热史	4.17
咳嗽史	3.55
SIRI	2.54
WBC	1.71
湿罗音	1.31
SII	1.24

## 2.2 单因素分析

本研究还通过 Logistic 回归对特征变量进行筛选。单因素分析发现,合并组患者年龄、地区、发热史、咳嗽史、咳嗽时间、湿罗音、WBC、NEU、LYM、

PLT、MONO、LDH、NLR、PLR、MLR、SII、SIRI 与肺炎组患者比较,差异有统计学意义( $P < 0.05$ ),如表 2 所示。

表 2 两组基线资料单因素分析

因素	合并组( $n=42$ )	肺炎组( $n=170$ )	$Z/t/\chi^2$	$P$
年龄/岁	11.5(4.3,15.0)	7.0(5.0,8.0)	3.063	<0.05
性别(男/女)	16/26	84/86	1.731	>0.05
地区(城市/农村)	8/34	159/11	111.739	<0.05
发热史(是/否)	19/23	154/16	46.140	<0.05
咳嗽史(是/否)	36/6	169/1	19.790	<0.05
咳嗽时间/d	26(10,57.5)	4(3,6)	6.341	<0.05
湿罗音(是/否)	10/32	83/87	8.557	<0.05
WBC/( $\times 10^9$ 个/L)	9.30(6.77,10.87)	6.72(5.49,8.42)	4.322	<0.05
NEU/%	66.09 $\pm$ 14.09	61.55(55.60,68.40)	2.600	<0.05
LYM/%	24.55 $\pm$ 14.02	29.38 $\pm$ 9.35	-2.122	<0.05
HGB/(g/L)	118.45 $\pm$ 19.01	123.00(116.00,128.00)	-0.931	>0.05
PLT/( $\times 10^9$ 个/L)	344.98 $\pm$ 114.68	249.00(206.00,296.00)	4.617	<0.05
MONO/( $\times 10^9$ 个/L)	0.74(0.54,0.87)	0.54(0.44,0.67)	4.353	<0.05
CRP/(mg/L)	18.25(7.55,38.22)	14.44(6.15,26.82)	1.226	>0.05
LDH/(U/L)	276.38 $\pm$ 80.50	318.00(266.00,375.00)	-3.726	<0.05
NLR	2.95(1.68,4.83)	2.14(1.53,3.01)	2.890	<0.05
PLR	13.75(8.72,28.57)	8.62(6.36,11.83)	4.598	<0.05
MLR	0.03(0.01,0.06)	0.01(0.01,0.02)	4.084	<0.05
SII	878.86(539.51,2165.14)	551.26(371.77,752.00)	4.231	<0.05
SIRI	2.19(1.03,5.28)	1.16(0.78,1.86)	3.846	<0.05

## 2.3 多因素 Logistic 回归分析

随后通过 ROC 曲线计算每个计量数据截值对数据进行赋值,将其转化为二分类数据,如表 3 所示。最后通过多因素 Logistic 回归分析发现,地区、发热史、咳嗽史、咳嗽时间、WBC 是儿童 TBTB 合并 MPP 的危险因素( $P < 0.05$ ),如表 4 所示。

表 3 差异因素赋值表

因素	赋值
年龄/岁	$\geq 10.50=1$ , $<10.50=0$
性别	男=1, 女=0
地区	城市=1, 农村=0
发热史	是=1, 否=0
咳嗽史	是=1, 否=0

续表

因素	赋值
咳嗽时间/d	$\geq 9.50 = 1$ , $< 9.50 = 0$
湿罗音(是/否)	是 = 1, 否 = 0
WBC/( $\times 10^9$ 个/L)	$\geq 8.44 = 1$ , $< 8.44 = 0$
NEU/%	$\geq 72.85 = 1$ , $< 72.85 = 0$
LYM/%	$\geq 18.10 = 1$ , $< 18.10 = 0$
PLT/( $\times 10^9$ 个/L)	$\geq 278.50 = 1$ , $< 278.50 = 0$
MONO/( $\times 10^9$ 个/L)	$\geq 0.68 = 1$ , $< 0.68 = 0$

续表

因素	赋值
LDH/(U/L)	$\geq 223.50 = 1$ , $< 223.50 = 0$
NLR	$\geq 3.51 = 1$ , $< 3.51 = 0$
PLR	$\geq 12.92 = 1$ , $< 12.92 = 0$
MLR	$\geq 0.04 = 1$ , $< 0.04 = 0$
SII	$\geq 792.58 = 1$ , $< 792.58 = 0$
SIRI	$\geq 2.38 = 1$ , $< 2.38 = 0$
分组	合并组 = 1, 肺炎组 = 0

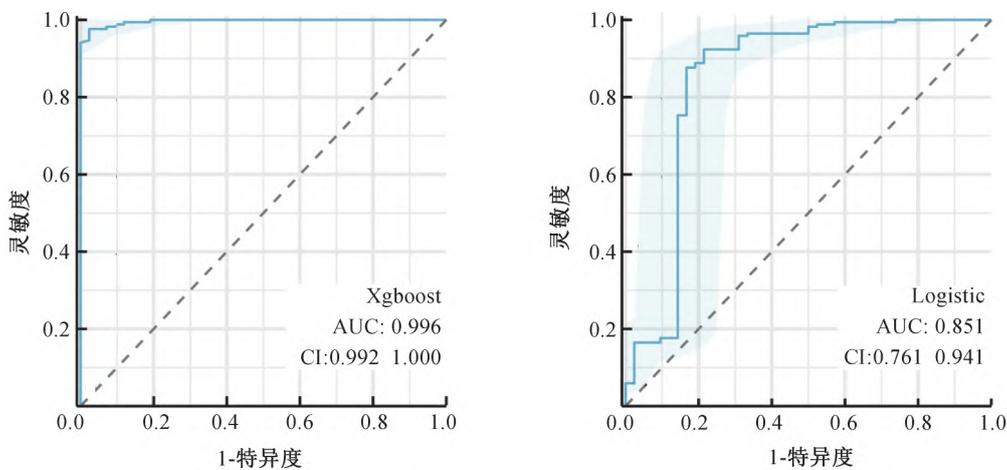
表 4 多因素 Logistic 回归分析

因素	$\beta$	SE	Wald $\chi^2$	P	OR	95%CI	
						下限	上限
地区	-6.593	1.639	16.181	<0.05	0.001	<0.001	0.048
发热史	-2.239	1.039	4.644	<0.05	0.107	0.020	0.861
咳嗽史	-7.922	3.701	4.581	<0.05	<0.001	<0.001	0.597
咳嗽时间	6.162	1.645	14.029	<0.05	8.893	4.313	13.543
WBC	3.576	1.248	8.211	<0.05	5.198	2.034	8.690

2.4 XGBoost 模型与 Logistic 回归模型能效评估

两个模型分别绘制 ROC 预测曲线,结果显示, XGBoost 的 AUC 为 0.996,而 Logistic 回归模型的 AUC 仅为 0.851,如图 1、表 5 所示。采用 Delong 检

验评估两个模型的效能,结果显示,XGBoost 模型预测效能要显著高于 Logistic 回归模型( $P < 0.05$ ),如表 6 所示。



注: A.XGBoost 模型; B.Logistic 回归模型。

图 1 XGBoost 模型与 Logistic 回归模型预测 TBTB 合并 MPP 患儿的 ROC 曲线

表5 ROC曲线详细参数

预测变量	AUC	95%CI	灵敏度/%	特异度/%	约登指数/%
XGBoost	0.996	0.992 - 1.000	97.64	97.61	95.26
Logistic	0.851	0.761 - 0.941	87.65	83.33	70.98

表6 Delong 检验评估两个模型效能差异

检验结果对	Z	P	AUC 差异	SE 差值	95%CI	
					下限	上限
XGBoost -Logistics	3.156	0.002	0.145	0.219	0.055	0.236

### 3 讨论

TBTB 合并 MPP 是临床中较为复杂且严重的呼吸道感染,尤其对儿童的健康影响深远。合并发生的 TBTB 与 MPP 不仅可能引起持续的咳嗽、发热等典型症状,还可能加重病情,对儿童的肺部健康造成长期损害<sup>[11]</sup>。由于这两种疾病在临床表现和影像学特征上存在一定的重叠,准确的诊断变得尤为关键。目前,临床诊断往往依赖多次影像学检查,但频繁的 X 光或 CT 检查可能给儿童带来额外的辐射暴露,增加健康风险<sup>[12]</sup>。因此,尽管影像学检查仍然是诊断 TBTB 合并 MPP 的重要手段,但如何减少误诊与漏诊,提升诊断准确率成为临床关注的重点。随着技术的进步,机器学习模型为复杂疾病的诊断提供了新的可能,能够辅助临床医生更精准地识别 TBTB 合并 MPP,优化治疗策略,改善患儿的预后。

在现代医学研究和临床诊断中,数据分析和机器学习技术的应用已经与传统模式显著不同<sup>[13]</sup>。XGBoost 模型和 Logistic 回归模型在临床医学中的应用展现了独特的价值。本研究比较了这两种模型在诊断 TBTB 合并 MPP 儿童中的效能。通过 ROC 曲线分析发现,XGBoost 的 AUC 为 0.996,而 Logistic 回归模型的 AUC 仅为 0.851,显示 XGBoost 模型的最佳预测效能。本研究推荐在诊断 TBTB 合并 MPP 患儿时使用 XGBoost 模型,以实现更准确和可靠的诊断效果。这种差异可能源于模型的复杂性、特征选择能力、过拟合控制、数据质量以及参数调优等因素的不同。由于 Logistic 回归模型相对较低的复杂性和缺乏特征选择机制,它在这种复杂的诊断任务中表现不佳。之前的一项研究<sup>[14]</sup>比较了 XGBoost 模型和

LASSO-Logistic 模型在预测患者急性肾损伤能力,发现 XGBoost 模型的约登指数显著高于 LASSO-Logistic 回归模型。同时,在肝癌和胆囊癌的预测中,XGBoost 模型的 AUC 值也较 Logistic 回归模型高<sup>[15]</sup>。这表明比较不同机器学习模型在临床预测和诊断任务中的重要性,通过比较可以选择最适合特定临床任务和数据集的模型。

应用 XGBoost 模型对患者数据进行分析,识别出与儿童 TBTB 合并 MPP 密切相关的八个特征变量。从生物学和临床医学的角度看,这些变量之间可能具有内在逻辑。农村地区由于医疗资源的缺乏和环境卫生条件较差,使得儿童更容易接触到结核杆菌和其他呼吸道病原体<sup>[16]</sup>。长时间的咳嗽可能反映呼吸道感染的持续和严重程度。虽然没有发热史,但持续的咳嗽可能是气管支气管结核的早期症状。较少的咳嗽史可能表明患儿对于呼吸道病原体的敏感性或易感性较高<sup>[17]</sup>。SIRI 和白细胞的升高可能是身体对感染的免疫反应,表明存在严重的炎症和感染。湿啰音的减少可能与病原体引起的呼吸道黏液减少或阻塞有关,这可能与支原体肺炎的病理机制相关<sup>[18]</sup>。SII 的升高也可能是感染和炎症的指标,进一步证实这种病理条件的严重性<sup>[19]</sup>。通过对这些变量及其与疾病的关联进行分析,为更好地理解 TBTB 合并 MPP 的病理机制和临床特征提供依据。尽管可能还需要更多研究来验证和深入理解这些变量与病症之间的具体关系,并探索其潜在的生物学机制。

本研究存在几个主要局限。首先,由于样本量相对较少,可能无法充分反映患儿群体的整体特征。

其次,由于研究设计为回顾性研究,存在一定的选择偏倚风险。另外,本研究仅比较了两种机器学习模型,实际上还有其他许多机器学习和统计模型可能同样适用于临床诊断任务。此外,本研究未能包括所有可能的危险因素和潜在的混杂变量,可能会对最终结果和结论产生影响。为得到更准确可靠的结果,未来的研究可以增加样本量,采用前瞻性设计,比较更多模型,并考虑更多危险因素。

#### 4 结论

XGBoost 模型在诊断 TBTB 合并 MPP 方面表现出色,优于 Logistic 回归模型。该模型成功识别了 8 个与疾病密切相关的特征变量,为理解病理机制提供了新视角。

#### 参考文献

- [1] 阮淑金,曾坚,陈敬芳,等.结核病治疗依从性状况、影响因素及干预措施研究进展[J].结核与肺部疾病杂志,2023,4(5):419-424.
- [2] 吴伟丽,丁军,张晓强.气管支气管结核患儿并发肺不张的影响因素分析[J].预防医学,2023,35(10):907-910.
- [3] 鞠春兰,王倩,刘玉琳,等.经支气管镜钬激光联合冷冻治疗儿童气管支气管结核的临床分析[J].激光杂志,2023,44(6):250-253.
- [4] 赵顺英,钱素云,陈志敏,等.儿童肺炎支原体肺炎诊疗指南(2023年版)[J].传染病信息,2023,36(4):291-297.
- [5] 宾松涛,胡晓琴,杨炳艳,等.儿童气管支气管结核九例临床特征及误诊分析[J].临床误诊误治,2021,34(4):19-23.
- [6] 谭敏,赵川,王文平.支气管结核六例误诊病例剖析[J].临床误诊误治,2018,31(6):7-9.
- [7] 高见,纪青.纤维支气管镜下局部注药对气管、支气管结核疾病的治疗效果[J].中国药物滥用防治杂志,2022,28(10):1412-1415.
- [8] 潘唯,吴琼华,陆兴.药物雾化吸入疗法联合纤支镜介入疗法治疗支气管内膜结核患者的疗效分析[J].系统医学,2022,7(19):19-22.
- [9] 周林,刘二勇,孟庆琳,等.《WS 288—2017 肺结核诊断》标准实施后肺结核诊断质量评估分析[J].中国防痨杂志,2020,42(9):910-915.
- [10] 王天有,申昆玲,沈颖.诸福棠实用儿科学[M].9版.北京:人民卫生出版社,2022.
- [11] WANG Y C, YU X R, LIU F, et al. Respiratory microbiota imbalance in children with *Mycoplasma pneumoniae* pneumonia[J]. Emerg Microbes Infect, 2023, 12(1): 2202272.
- [12] BRAMBILLA M, FRUSH D P, REHANI M M. Cumulative radiation dose from medical imaging in children with non-cancerous disease[J]. J Am Coll Radiol, 2020, 17(12): 1547-1548.
- [13] YU X, JI Y W, HUANG M J, et al. Machine learning for acute kidney injury: changing the traditional disease prediction mode[J]. Front Med, 2023, 10: 1050255.
- [14] ZHANG Y L, WANG Y M, XU J R, et al. Comparison of prediction models for acute kidney injury among patients with hepatobiliary malignancies based on XGBoost and LASSO-Logistic algorithms[J]. Int J Gen Med, 2021, 14: 1325-1335.
- [15] HUANG J C, TSAI Y C, WU P Y, et al. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method[J]. Comput Methods Programs Biomed, 2020, 195: 105536.
- [16] PINTO P F P S, TEIXEIRA C S S, ICHIHARA M Y, et al. Incidence and risk factors of tuberculosis among 420854 household contacts of patients with tuberculosis in the 100 Million Brazilian Cohort (2004-18): a cohort study[J]. Lancet Infect Dis, 2024, 24(1): 46-56.
- [17] KIM C, KO Y, MOON J W, et al. Incidence, risk factors, and final causes for misdiagnosis of tuberculosis in the Republic of Korea: a population-based longitudinal analysis[J]. Eur Respir J, 2022, 60(4): 2201461.
- [18] VENKATAPPA T, SHEN D, AYALA A, et al. Association of *Mycobacterium tuberculosis* infection test Results with risk factors for tuberculosis transmission[J]. J Clin Tuberc Other Mycobact Dis, 2023, 33: 100386.
- [19] VENKATAPPA T, SHEN D, AYALA A, et al. Association of *Mycobacterium tuberculosis* infection test results with risk factors for tuberculosis transmission[J]. J Clin Tuberc Other Mycobact Dis, 2023, 33: 100386.

[收稿日期: 2024-01-08]

[责任编辑: 桂根浩 英文编辑: 周寿红]