临床研究 DOI：10.13406/j.cnki.cyxb.003962

基于机器学习算法构建HIV/AIDS合并马尔尼菲篮状菌感染诊断模型的探讨

谢雨伶，杨　坤，但文冉
（重庆市公共卫生医疗救治中心，重庆　400016）

【摘　要】**目的**：基于机器学习算法构建人类免疫缺陷病毒/获得性免疫缺陷综合征（human immunodeficiency virus and acquired immunodeficiency syndrome，HIV/AIDS）患者合并马尔尼菲篮状菌（*Talaromyces marneffei*，TM）感染的诊断模型，以实现辅助早期诊断和提升诊断灵敏度。**方法**：回顾性收集2020年1月至2023年9月在重庆市公共卫生医疗救治中心共201例HIV/AIDS-Mp1p抗原阳性患者实验室数据，筛选得到确诊TM（TM组）91例和未感染TM（非TM组）110例。将数据通过统计学分析获得2组间差异性指标，并构建logistic回归、随机森林分类器、决策树分类器模型。再通过最小绝对收缩和选择算子（least absolute shrinkage and selection operator，Lasso）回归筛选差异变量，并构建Lasso-logistic回归、随机森林和决策树分类器模型。分析所有模型筛选的高贡献特征指标，采用正确率、精确率、受试者工作特征（receiver operating characteristic，ROC）曲线和曲线下面积（area under the curve，AUC）评价模型的诊断性能。**结果**：通过TM组与非TM组检验指标比较，删除了白细胞计数和性别2个差异无统计学意义的指标（$P>0.05$），纳入红细胞计数、血小板计数、血红蛋白含量、C-反应蛋白、白细胞介素-6、降钙素原、年龄、$CD4^+$T淋巴细胞计数、HIV-RNA含量、（1-3）-β-D葡聚糖检测、曲霉半乳甘露聚糖抗原检测11个差异有统计学意义（$P<0.05$）的指标构建logistic回归模型、决策树和随机森林分类模型。进一步通过Lasso回归分析删除了$CD4^+$T淋巴细胞计数和红细胞计数指标，剩余9个变量纳入构建了Lasso-logistic回归、决策树和随机森林分类模型，其AUC均大于纳入11个变量构建的同种模型，其中随机森林分类模型（$n=9$）的诊断性能最佳，正确率为0.797、精确率为0.794、AUC=0.822（95%CI=0.719~0.924）。**结论**：在不同诊断模型中实验室检验指标特征重要性不同，经Lasso回归筛选变量后再构建模型能提高诊断性能，构建的所有模型中随机森林分类模型诊断性能最好。基于机器学习算法和临床检验数据建立诊断模型，有利于辅助临床早期诊断HIV/AIDS合并TM感染。

【关键词】机器学习；马尔尼菲篮状菌；人类免疫缺陷病毒；分类器诊断模型
【中图分类号】R756　　【文献标志码】A

Establishment of a diagnostic model for human immunodeficiency virus/acquired immunodeficiency syndrome co-infection with *Talaromyces marneffei* based on machine learning algorithms

Xie Yuling，*Yang Kun*，*Dan Wenran*
（*Chongqing Public Health Medical Treatment Center*）

【**Abstract**】***Objective***：To establish a diagnostic model for human immunodeficiency virus/acquired immunodeficiency syndrome（HIV/AIDS）patients co-infected with *Talaromyces marneffei*（TM）using machine learning algorithms，and to help realize early diagnosis and improve diagnostic sensitivity. ***Methods***：A retrospective analysis was performed for the laboratory test data of 201 HIV/AIDS patients with positive Mp1p antigen from January 2020 to September 2023，among whom 91 were diagnosed with TM infection（TM group）and 110 were not infected with TM（non-TM group）. The statistical analysis was performed to identify differential indicators between the two groups，and the models of logistic regression，Random Forest Classifier，and Decision Tree Classifier were constructed. The Least absolute shrinkage and selection operator（LASSO）regression was used to identify differential variables，and then the models of LASSO-logistic regression，Random Forest Classifier，and Decision Tree Classifier were constructed. The characteristic indicators with high contribution identified by all models were analyzed，and the diagnostic performance of the model was assessed by accuracy，precision，the receiver operating characteristic（ROC）curve，and the area under the ROC curve（AUC）. ***Results***：Through

作者介绍：谢雨伶，Email：lxd336@163.com，研究方向：医学检验。
通信作者：但文冉，Email：danwenran@163.com。
基金项目：重庆市沙坪坝区科卫联合医学科研资助项目（编号：2024SQKWLHMS043）；重庆市科技局自然科学基金面上资助项目（编号：CSTB2023NSCQ-MSX1060）；吴阶平医学基金会科研专项基金资助项目（编号：320.6750）。
优先出版：https://link.cnki.net/urlid/50.1046.R.20251016.1319.002　2025-10-16

comparison of test indicators between the TM group and the non-TM group, two indicators without statistical significance (leukocyte count and sex) were excluded (P>0.05), and eleven indicators with statistical significance were used to establish the models of logistic regression, Decision Tree Classifier, and Random Forest Classifier, i.e., red blood cell count, platelet count, hemoglobin content, C-reactive protein, interleukin-6, procalcitonin, age, $CD4^+$ T lymphocyte count, HIV-RNA level, (1-3)-β-D-glucan test, and *Aspergillus galactomannan* antigen test (P<0.05). The indicators of $CD4^+$ T lymphocyte count and red blood cell count were removed by the LASSO regression analysis, and the remaining nine variables were included to establish the models of LASSO-logistic regression, Decision Tree Classifier, and Random Forest Classifier, which had a greater AUC than the models established based on the eleven variables; the Random Forest Classifier model (n=9) had the best diagnostic performance, with an accuracy rate of 0.797, a precision rate of 0.794, and an AUC of 0.822 (95%CI=0.719-0.924). ***Conclusion***: Laboratory test indicators have different levels of importance in different diagnostic models, and the model established after screening for variables based on LASSO regression can improve the diagnostic performance of the model. The Random Forest Classifier model has the best diagnostic performance among all the models established. The establishment of diagnostic model based on machine learning algorithms and clinical test data can help with the early diagnosis of HIV/AIDS with TM infection in clinical practice.

【Key words】machine learning; *Talaromyces marneffei*; human immunodeficiency virus; classifier diagnostic model

马尔尼菲篮状菌（*Talaromyces marneffei*，TM）是一种温度双相型机会性致病真菌，主要在东南亚地区和中国南方流行[1]，免疫功能正常的宿主可以携带TM，但当免疫系统受到抑制或损害时，就会出现致命的散播型真菌感染，尤其是在获得性免疫缺陷综合征（acquired immunodeficiency syndrome，AIDS）患者中[2]。马尔尼菲篮状菌病已从一种罕见的感染激增为继结核病、口腔念珠菌感染和肺孢子菌肺炎之后，最常见的人类免疫缺陷病毒（human immunodeficiency virus，HIV）患者中的机会性感染，也是HIV相关死亡的主要原因，超过结核病和隐球菌脑膜炎[1-3]。

TM感染可通过侵入与单核-巨噬细胞网状内皮系统相关的器官和部位而分为散播型和局灶型，没有及时诊断和有效的抗真菌治疗的TM病患者病死率可达到90%，及时有效的诊断则可将病死率降到30%[4]。TM感染的早期诊断是影响HIV/AIDS患者预后和生存结局的重要因素，而组织或体液的真菌学培养阳性是TM感染的“金标准”，可该方法不仅耗时长（7~14 d），而且灵敏度较低（血培养阳性率76%左右）[5]，并会遗憾地错过治疗时间窗口期。因此迫切需要早期诊断方法进行及时治疗干预。

目前，机器学习的各种算法在患者数据集中显示出潜在的应用前景，通过深入挖掘变量特征建立疾病诊断模型辅助诊断。本文将通过机器学习算法区分HIV/AIDS患者中TM与非TM的临床实验室检验的特征指标，构建logistic回归模型和分类器模型进行TM诊断并分析模型诊断性能。

1 资料与方法

1.1 研究对象及纳入标准

回顾性收集了2020年1月至2023年9月在重庆市公共卫生医疗救治中心有完整实验室检验资料的HIV/AIDS住院患者电子病历资料共201例，收集电子病例数据内容包括HIV/AIDS住院患者的一般人口统计学、常规实验室检验指标白细胞计数（white blood cell count，WBC）、红细胞计数（red blood cell count，RBC）、血小板计数（platelet count，PLT）、C-反应蛋白（C-reactive protein，CRP）、血红蛋白含量（hemoglobin content，Hb）、降钙素原（procalcitonin，PCT）、白细胞介素-6（interleukin-6，IL-6）、$CD4^+$T细胞计数、人类免疫缺陷病毒核糖核酸（human immunodeficiency virus ribonucleic acid，HIV-RNA）含量、真菌学检验结果[(1-3)-β-D葡聚糖检测（G试验）、曲霉半乳甘露聚糖抗原检测（GM试验）、马尔尼菲抗原荧光检测（TM-Mp1p-Ag）]以及真菌培养结果，汇总整理成表格后，用于统计分析和构建机器学习预测模型。从血液和（或）痰液、肺泡灌洗液、组织培养出TM即为确诊马尔尼菲篮状菌感染。

本研究的数据纳入标准为：①纳入患者均为经确诊为HIV/AIDS阳性患者；②血清TM抗原检测阳性；③确诊TM感染或非TM感染。排除标准为：①无完整实验室检验指标数据；②合并其他霉菌感染。

1.2 统计学方法

首先对TM组和非TM组的实验室检查指标使用GraphPad Prism 10软件分析数据，采用Shapiro-Wilk（S-W）检验进行正态性检验，P<0.05表明不服从正态分布，非正态分布的计量资料以中位数（四分位间距）[M_d(P_{25},P_{75})]描述，组间比较采用Wilcoxon秩和检验；S-W检验P>0.05表明服从正态分布，计量资料以均数±标准差（$\bar{x}\pm s$）表示，2组间比较采用两独立样本t检验；计数资料采用频数或百分率表示，多组间比较采用卡方检验。检验水准α=0.05。

1.3 Lasso回归筛选TM发生的预测因素

以患者是否发生TM（临床标本培养TM阳性）为因变量，将1.2中筛选出差异有统计学意义的指标作为自变量按表1进行赋值，采用R软件中glmnet进行Lasso分析，根据惩罚系数λ筛选出有意义的预测因素。

表1 Lasso回归赋值方法

变量	赋值
因变量	
确诊TM感染	否=0，是=1
自变量	
年龄	连续变量，原值代入
CRP	连续变量，原值代入
HIV-RNA	连续变量，原值代入
PLT	连续变量，原值代入
PCT	连续变量，原值代入
IL-6	连续变量，原值代入
$CD4^+$T细胞计数	等级变量1=<100个/μL，2=100~350个/μL，3=>350个/μL
Hb	连续变量，原值代入
RBC	连续变量，原值代入
G试验	阴性=0，阳性=1
GM试验	阴性=0，阳性=1

1.4 logistic回归分析

对TM组和非TM组患者的临床资料进行比较分析后，$P<0.05$的临床检验指标被纳入构建logistic回归模型。获得$P<0.05$的有意义特征变量。再根据Lasso回归筛选的有意义变量构建Lasso-logistic回归模型。R4.4.1软件glmnet包函数用于构建logistic回归模型

1.5 分类器模型构建

通过收集的临床资料数据集构建决策树和随机森林分类器模型以评估HIV/AIDS患者是否感染TM。数据集按照3∶1随机分为训练集和测试集，分别使用全部数据预处理得到的11个变量构建分类器模型，和使用经Lasso回归筛选的变量构建分类器模型。R4.4.1软件rpart包函数用于构建决策树，randomForest包函数用于构建随机森林。

1.6 回归模型和分类器模型的评价

使用受试者工作特征（receiver operating characteristic，ROC）曲线、曲线下面积（area under the curve，AUC）、精确率、正确率评估模型诊断性能，通过三折交叉验证对正确率和精确率进行平均。ROC曲线和AUC是通过R4.4.1软件中pROC构建并计算的。精确率=真阳性值/预测为阳性的个数、正确率=（真阳性值+真阴性值）/总个数。

2 结 果

2.1 2组患者基本临床资料比较

201例确诊HIV/AIDS合并TM抗原阳性患者电子病例数据被纳入研究，临床体液或血液标本培养出TM组91例，非TM组患者110例。与非TM组相比，TM组年龄较小（$P<0.05$），RBC、Hb、PLT较低（$P<0.05$），CRP、PCT、IL-6、$CD4^+$T细胞计数、HIV-RNA含量、G试验阳性率和GM试验阳性率均较高（$P<0.05$）。2组性别和WBC差异无统计学意义（$P>0.05$）。见表2。

表2 2组患者基本临床资料比较[n,%;$M_d(P_{25},P_{75})$;$\bar{x}\pm s$]

因素	TM组（n=91）	非TM组（n=110）	t/χ^2值	P值
性别			1.739	0.187
女	30(33)	27(25)		
男	61(67)	83(75)		
年龄（岁）	46(35,51)	52(36,62.5)	14.530	<0.05
白细胞计数（$\times10^9$/L）	3.43(2.21,2.58)	4.23(2.78,6.20)	7.297	0.063
红细胞计数（$\times10^{12}$/L）	2.98±0.66	3.37±0.80	3.720	<0.01
血红蛋白（g/L）	85.74±21.86	102.5±26.32	4.830	<0.01
血小板计数（$\times10^9$/L）	92(37,170)	146(9.83,87.83)	14.640	<0.05
C-反应蛋白（mg/L）	71.45(28.37,110.40)	35.45(17.77,91.08)	12.030	<0.05
降钙素原（ng/mL）	1.11(0.31,3.97)	0.19(0.08,0.86)	33.717	<0.01
IL-6（pg/mL）	52.2(20.00,120.40)	25.82(7.90,65.46)	13.190	<0.05
$CD4^+$T细胞计数（个/μL）	19(11,33)	45.5(19.75,119.00)	83.270	<0.01
HIV-RNA（拷贝/μL）	714 000(44 500,2 940 000)	214 000(57.8,919 250)	19.830	<0.001
G试验			34.000	<0.001
阴性	19(21)	68(62)		
阳性	72(79)	42(38)		
GM试验			45.550	<0.001
阴性	9(10)	61(55)		
阳性	82(90)	49(45)		

2.2 logistic回归分析TM发生的影响因素

将TM组与非TM组有明显差异的实验室检验指标进行logistic回归分析，以筛选TM发生的影响因素，因变量：是否确诊感染TM。首先将所有差异指标（n=11）纳入分析，通过逐步回归分析后仅保留明显性指标GM试验和G试验阳性进行统计建模，并存在明显的线性关系（P<0.01）（表3）。为降低逐步回归的过拟合，进一步进行二分类logistic回归，结果显示（表4），G试验阳性、GM试验阳性对TM感染有明显影响（P<0.05），G试验阳性使TM培养阳性概率高了152.21%，GM试验阳性使TM培养阳性概率高了317.52%。其余指标则对确诊TM感染的结果不产生显著影响（P>0.05）。之后通过Lasso回归分析和交叉验证确认最佳参数λ=0.04筛选出IL-6、年龄、HIV-RNA、CRP、PCT、PLT、Hb、G试验阳性、GM试验阳性9个变量为预测变量，并构建Lasso-logistic回归模型（表5），模型提示对TM发生的影响因素主要是GM和G试验阳性（P<0.05）。根据Lasso-logistic回归模型，G试验阳性使TM培养阳性概率高了161.4%，GM试验阳性使TM培养阳性概率高了368.01%。

2.3 基于机器学习的TM诊断模型构建

按照3∶1的比例将数据分为训练集和测试集，分别构建决策树分类模型和随机森林分类模型。首先将2.1筛选的TM组与非TM组有明显差异的实验室检验指标（n=11）全部纳入模型构建，通过决策树分类模型计算出诊断TM的主要特征指标为GM试验阳性（20.40%）、PCT（12.60%）、HIV-

表3 逐步回归模型结果

项目	β	Se	t	P	VIF	R^2	F
常数	0.075	0.054	1.403	0.162			
GM试验阳性	0.377	0.073	5.133	<0.01	1.333	0.267	F=36.012，P<0.01
G试验阳性	0.232	0.071	3.291	<0.01	1.333		

表4 二分类logistic回归结果

TM组	β	Se	$Wald$	P	OR	OR（95%CI）
常数	−0.372	1.139	0.107	0.744	0.689	0.074~6.423
IL-6	0.000	0.000	0.043	0.835	1.000	1.000~1.000
年龄	−0.007	0.013	0.271	0.603	0.993	0.968~1.019
G试验阳性	0.925	0.408	5.130	0.024	2.522	1.133~5.616
HIV-RNA	0.000	0.000	2.603	0.107	1.000	1.000~1.000
GM试验阳性	1.429	0.478	8.940	0.003	4.175	1.636~10.655
Hb	−0.021	0.021	0.990	0.320	0.979	0.940~1.021
CRP	0.001	0.002	0.222	0.637	1.001	0.997~1.004
PCT	0.003	0.006	0.172	0.679	1.003	0.990~1.015
PLT	−0.001	0.002	0.350	0.554	0.999	0.995~1.003
RBC	0.290	0.664	0.191	0.662	1.336	0.364~4.909
$CD4^+$T淋巴细胞	−0.002	0.002	1.023	0.312	0.998	0.993~1.002

表5 Lasso-logistic回归结果

TM组	β	Se	$Wald$	P	OR	OR（95%CI）
常数	−0.217	1.111	0.038	0.845	0.805	0.091~7.101
年龄	−0.008	0.013	0.406	0.524	0.992	0.967~1.017
CRP	0.001	0.002	0.178	0.673	1.001	0.997~1.004
Hb	−0.014	0.008	3.186	0.074	0.986	0.971~1.001
PLT	−0.001	0.002	0.366	0.545	0.999	0.995~1.003
PCT	0.003	0.006	0.250	0.617	1.003	0.991~1.015
IL-6	0.000	0.000	0.041	0.840	1.000	1.000~1.000
HIV-RNA	0.000	0.000	0.912	0.340	1.000	1.000~1.000
GM试验阳性	1.543	0.468	10.868	0.001	4.680	1.870~11.715
G试验阳性	0.961	0.399	5.792	0.016	2.614	1.195~5.717

RNA 含量（19.90%）、Hb 含量（9.70%）和 $CD4^+$T 细胞计数（9.40%），通过随机森林分类模型计算出的诊断 TM 的主要特征指标为 PCT（11.90%）、PLT（10.10%）、HIV-RNA 含量（12.00%）、Hb 含量（9.20%）和 $CD4^+$T 细胞计数（10.00%）。见表 6。将 Lasso 回归筛选的 9 个变量构建分类模型，通过决策树分类模型计算出诊断 TM 的主要特征指标为 PCT（28.40%）、GM 试验阳性（20.4%）、Hb 含量（15.40%）、HIV-RNA（10.10%）、年龄（10.10%），通过随机森林分类模型计算出的诊断 TM 的主要特征指标为 PCT（14.20%）、Hb 含量（13.60%）、HIV-RNA（12.40%）、PLT（12.40%）。见表 7。

表 6　机器学习分类模型特征重要性（n=11，%）

特征	特征重要性	
	决策树分类	随机森林分类
年龄	3.90	8.60
$CD4^+$T 淋巴细胞	9.40	10.00
CRP	3.70	6.30
PLT	6.80	10.10
GM 试验阳性	20.40	6.80
Hb	9.70	9.20
G 试验阳性	2.20	7.50
RBC	11.90	8.90
HIV-RNA	19.90	12.00
IL-6	0.00	8.60
PCT	12.60	11.90

表 7　机器学习分类模型特征重要性（n=9，%）

特征	特征重要性	
	决策树分类	随机森林分类
年龄	10.10	11.40
Hb	15.40	13.60
IL-6	5.40	11.40
CRP	3.70	7.60
HIV-RNA	10.10	12.40
PLT	3.50	12.40
G 试验阳性	2.90	6.90
PCT	28.40	14.20
GM 试验阳性	20.40	10.10

2.4　诊断模型评价

根据混淆矩阵和 ROC 曲线下面积评估诊断模型性能，通过三折交叉验证对正确率和精确率进行平均后对 2 种机器学习分类模型进行评价，根据 ROC AUC 对 6 种诊断模型进行综合评价。根据混淆矩阵（图 1）和 ROC 曲线（图 2）结果，如表 8 所示，11 变量分类模型和 9 变量分类模型中，随机森林分类模型正确率和精确率均优于决策树分类模型，Lasso-logistic 回归诊断性能优于 logistic 回归，构建的 6 种诊断模型中，随机森林分类器（n=9）模型的 AUC 最大，具有最佳诊断性能。

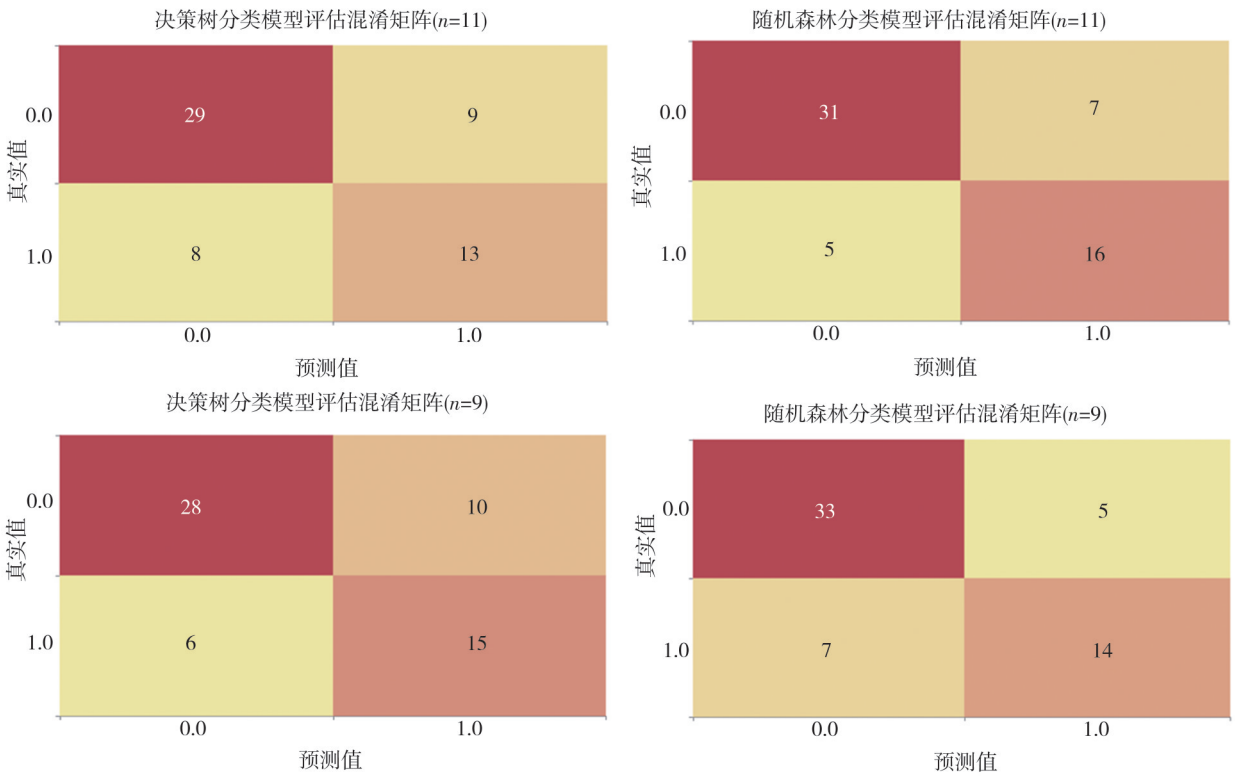


图 1　机器学习分类器模型混淆矩阵

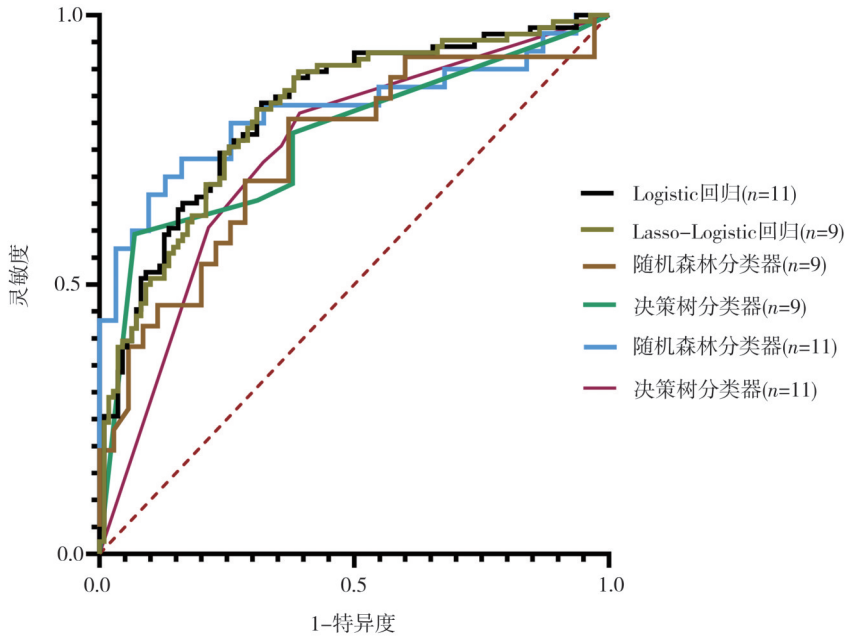


图2 6种诊断模型ROC曲线

表8 诊断模型评价

模型	正确率	精确率	AUC(95%CI)
logistic回归(n=11)	0.745	0.749	0.737(0.607~0.869)
Lasso-logistic回归(n=9)	0.745	0.750	0.792(0.673~0.911)
决策树分类器(n=11)	0.712	0.715	0.743(0.623~0.863)
随机森林分类器(n=11)	0.763	0.759	0.820(0.707~0.933)
决策树分类器(n=9)	0.729	0.744	0.768(0.652~0.885)
随机森林分类器(n=9)	0.797	0.794	0.822(0.719~0.924)

3 讨 论

TM是HIV/AIDS患者的主要机会性感染病原体，作为胞内侵袭性真菌，TM主要通过侵犯单核-巨噬细胞网状内皮系统进入血液，然后经淋巴系统和血液循环扩散到相关器官和部位导致播散型TM病，其起病隐匿、临床表现复杂性和病死率高[6]。延迟诊断是TM病高病死率的重要原因，然而其明确诊断依赖于TM的组织病理学或培养证据，其因耗时较长和需要进行侵入性检查而在临床实践中存在明显的局限性。

当前已有多种实验手段用于早期辅助诊断TM，G试验和GM试验都被用于辅助诊断侵袭性真菌感染，但是GM广泛用于协助曲霉菌的诊断，G试验用于侵袭性真菌的大范围协助诊断，用于TM检测时则因为交叉反应而限制其临床应用。Mp1p抗原是存在于TM细胞壁中的特异性抗原，2022年全球地方性真菌病指南推荐进行Mp1p抗原检测作为TM筛选手段[7]。本研究以TM培养结果为金标准，收集的201例Mp1p抗原阳性HIV/AIDS患者确诊TM阳性率为45.27%(91/201)，表明基于Mp1p抗原阳性辅助其他实验室指标有助于提高TM病的早期诊断。

本研究通过回顾性分析HIV/AIDS-Mp1p阳性患者一般人口学特征和实验室检验指标，包括年龄、性别、G试验、GM试验、WBC、RBC、PLT、Hb、CRP、IL-6、PCT、$CD4^+$T淋巴细胞计数、HIV-RNA含量、TM培养结果，删除了无组间差异性指标性别和WBC，再以TM组/非TM组为结局，通过多因素logistic回归筛选独立危险因素，结果表明Mp1p抗原阳性联合G试验和GM试验阳性能明显提高诊断性能。

机器学习在医学诊断模型开发中已有充分研

究，本研究经过整合数据和筛选风险变量构建了决策树和随机森林2种机器学习分类器模型，2种模型基于对不同指标的重要性占比不同实现对TM结局的分类，不同指标虽然占比不同，但是都将HIV-RNA含量、Hb和PCT作为较重要的参数，这与既往的研究[8-9]一致，与炎症指标IL-6、CRP相比，PCT在真菌感染中似乎是更可靠的指标，比起WBC、RBC、PLT，Hb在TM感染中似乎更能反映骨髓抑制。

Lasso回归以λ正则项筛选变量，直至λ值增大，部分自变量回归系数收缩至0，并可从多重共线性变量集中选择以筛选预测变量[10]。本研究通过Lasso回归筛选删除了RBC、$CD4^+$T淋巴细胞计数，剩余的9个变量构建logistic回归分析、随机森林和决策树模型。根据AUC结果显示Lasso-logistic回归诊断性能优于logistic回归，而且经Lasso回归筛选的9个变量的决策树和随机森林分类模型比11个变量同种分类器AUC更大，表明诊断性能更高，而且在所有模型中随机森林分类模型（n=9）诊断性能最好。研究结果表明，随机森林分类模型能够通过纳入更少的变量达到相似的诊断效果，而且诊断性能更稳定。

本研究是基于Mp1p抗原阳性HIV/AIDS患者合并TM的诊断模型建立的探讨，由于Mp1p抗原检测结果的准确性依赖于检测技术和平台的可靠性，纳入其他实验室检验指标进行整合能提高诊断性能和弱化个体差异。同时通过对不同实验指标的回归筛选，优化诊断模型，并进一步分析不同模型的诊断价值。但是本研究尚存在一定的局限性，样本量的不足影响了模型的稳定性和泛化能力，只纳入了部分炎症相关和免疫相关指标的模型具有一定的偏倚，如果加入临床表现、抗生素使用等情况可能会更好拟合模型。本研究也提供了临床诊断TM的检验方向，可以依据现有的检验结果选择适合的模型进行分析，为早期诊断和及时用药提供一些技术手段。

参 考 文 献

[1] Cen JM, Huang J, Zeng W, et al. Clinical features of HIV positive talaromycosis marneffei patients and development of a risk prediction model[J]. Heliyon, 2023, 9(9): e20068.

[2] Zhang ZH, Li BP, Chai ZL, et al. Evolution of the ability to evade host innate immune defense by *Talaromyces marneffei*[J]. Int J Biol Macromol, 2023, 253(Pt 8): 127597.

[3] Son VT, Khue PM, Strobel M. Penicilliosis and AIDS in Haiphong, Vietnam: evolution and predictive factors of death[J]. Med Mal Infect, 2014, 44(11/12): 495-501.

[4] Wang F, Han RH, Chen S. An overlooked and underrated endemic mycosis-talaromycosis and the pathogenic fungus *Talaromyces marneffei*[J]. Clin Microbiol Rev, 2023, 36(1): e0005122.

[5] 李玥琪. 机器学习模型在HIV合并马尔尼菲篮状菌早期诊断中的应用研究[D]. 南宁：广西医科大学，2021.
Li YQ. Application of machine learning model in early diagnosis of HIV complicated with basilaria marneffei[D]. Nanning: Guangxi Medical University, 2021.

[6] Peng L, Shi Y, Zheng L, et al. Clinical features of patients with talaromycosis marneffei and microbiological characteristics of the causative strains[J]. Journal of Clinical Laboratory Analysis, 2022, 36(11): e24737.

[7] Chen XM, Ou X, Wang HD, et al. *Talaromyces marneffei* Mp1p antigen detection may play an important role in the early diagnosis of talaromycosis in patients with acquired immunodeficiency syndrome[J]. Mycopathologia, 2022, 187(2/3): 205-215.

[8] Wei HY, Liang WJ, Li B, et al. Clinical characteristics and risk factors of *Talaromyces marneffei* infection in human immunodeficiency virus-negative patients: a retrospective observational study[J]. World J Emerg Med, 2021, 12(4): 281-286.

[9] Jiang J, Meng S, Huang S, et al. Effects of *Talaromyces marneffei* infection on mortality of HIV/AIDS patients in Southern China: a retrospective cohort study[J]. Clin Microbiol Infect, 2019, 25(2): 233-241.

[10] 杜 然，滕 腾，赵云凤，等. Lasso-Logistic回归模型拟合临床因素、NF-κB/NLRP3信号通路预测心肌梗死后缺血性心肌病价值[J]. 中国急救复苏与灾害医学杂志，2024，19(6)：1015-1023.
Du R, Teng T, Zhao YF, et al. Lasso-Logistic regression model fitting clinical factors and NF-κB/NLRP3 predictive value of ischemic cardiomyopathy after myocardial infarction[J]. China J Emerg Resusc Disaster Med, 2024, 19(6): 1015-1023.

（收稿：2024-08-28；修回：2024-10-31；录用：2024-12-23）

（责任编辑：周一青）

本文引用格式：

谢雨伶，杨 坤，但文冉. 基于机器学习算法构建HIV/AIDS合并马尔尼菲篮状菌感染诊断模型的探讨[J]. 重庆医科大学学报，2025，50(12)：1744-1750.