文章编号:1671-7554(2024)02-0051-09 DOI:10.6040/j.issn.1671-7554.0.2023.1092

基于多检验变量和机器学习算法的结肠癌诊断模型建立及价值评估

梁永媛[1],蔡培飞[2],郑桂喜[1]
(1. 山东大学齐鲁医院检验科, 山东 济南 250012;2. 济宁医学院附属医院输血科,山东 济宁 272000)

摘要:目的 采用不同机器学习算法,建立基于多检验变量的结肠癌诊断模型,并评估其临床应用价值。**方法** 收集119例结肠癌患者(结肠癌组)和125例健康对照(健康对照组)的血清样本,提取血清外泌体,采用RT-qPCR方法测定miR-214-3p分子在两组中的表达水平,进而绘制受试者工作特征(receiver operating characteristic, ROC)曲线,评估其对结肠癌的诊断效能。同时,收集结肠癌组和健康对照组的常规检验项目结果。将以上指标均纳入研究筛选出特征性变量,采用11种不同算法结合ROC曲线和机器学习曲线综合评价筛选出最优算法,建立结肠癌诊断模型。**结果** 结肠癌组血清外泌体中miR-214-3p的表达水平明显高于健康对照组($P<0.001$),其诊断结肠癌的ROC曲线下面积(area under curve, AUC)为0.820,具有较好的诊断效能。将结肠癌组和健康对照组的血清外泌体miR-214-3p及30种常规检验指标纳入后,筛选出尿素、癌胚抗原、单核细胞计数、外泌体miR-214-3p共4个特征性变量,且逻辑回归算法是建立机器学习模型的最优算法,其AUC为0.93,且学习曲线呈现很好的拟合状态。**结论** 血清外泌体miR-214-3p是结肠癌的潜在标志物,基于4个特征性变量和逻辑回归算法建立的机器学习模型对结肠癌有良好的诊断效能。

关键词:结肠癌;miR-214-3p;外泌体;机器学习
中图分类号:R783 **文献标志码:**A

Establishment and value assessment of colon cancer diagnostic models based on multiple variables and different machine learning algorithms

LIANG Yongyuan[1], CAI Peifei[2], ZHENG Guixi[1]
(1. Department of Clinical Laboratory, Qilu Hospital of Shandong University, Jinan 250012, Shandong, China;
2. Department of Blood Transfusion, Affiliated Hospital of Jining Medical College, Jining 272000, Shandong, China)

Abstract: Objective To establish a colon cancer diagnostic model based on multiple variables using various machine learning algorithms and to assess its clinical application value. **Methods** Serum samples from 119 colon cancer patients and 125 healthy controls were collected. Serum exosome was extracted, and miRNA 214-3p (miR-214-3p) level was measured using RT-qPCR. Receiver operating characteristic (ROC) curve was plotted to evaluate the diagnostic efficiency of colon cancer. Additionally, 30 routine laboratory items of colon cancer patients and healthy controls were collected. Characteristic variables were screened, and 11 algorithms were used to establish the diagnostic model. The optimal model was selected with ROC and machine learning curves. **Results** The expression level of miR-214-3p in colon cancer patients was significantly higher than that in healthy controls ($P<0.001$), with the area under the ROC curve (AUC) being 0.820, indicating good diagnostic performance. After the expression level of miR-214-3p and other 30 routine laboratory items were enrolled, 4 characteristic variables were screened to establish the diagnostic model, including UREA, carcinoembryonic antigen, monocyte and miR-214-3p. The Logistic regression algorithm was identified as the optimal one (AUC=0.93). **Conclusion** Serum exosome miR-214-3p is a potential biomarker of colon cancer. The model based on 4 characteristic variables and Logistic regression algorithm has an excellent diagnostic performance for diagnosing colon cancer.

Key words: Colon cancer; miR-214-3p; Exosome; Machine learning

收稿日期:2023-12-05
基金项目:国家自然科学基金(82172347);山东省科技创新重大专项(2021CXGC010603)
通信作者:郑桂喜。E-mail:zhengg@ sdu.edu.cn

结肠癌已成为我国消化系统疾病中发病率第二位、患病率第一位的恶性肿瘤，其发病率以每年4%~8%的速度递增，且呈现低龄化趋势[1]。结肠癌患者早期临床症状隐匿，约60%患者就诊时已处于中晚期，错过了手术治愈的最佳时机，预后较差。近年来，结肠癌的筛查和治疗策略已取得较大进展，但纵观全球尤其是发展中国家，结肠癌的高死亡率仍未取得显著改善[2]。因此，寻找新型分子标志物、构建基于多指标联合检测的诊断模型，实现结肠癌的早诊早治是当前研究关注的焦点[3]。

研究表明，微小核糖核酸（micro ribonucleic acid，miRNA）在血清外泌体中稳定存在，作为肿瘤标志物具有高敏感度和特异性的优势。血清外泌体中miRNA分子已被发现在多种恶性肿瘤中具有较好的诊断价值[4-5]。随着实验室信息系统的发展，研究者越来越关注患者检验数据，尤其是探索大量检验数据间的相互关系和联合应用。既往国内外报道的基于海量数据模型的建立已有10余种不同算法，能够将纳入指标信息充分发掘，并在监督或非监督情况下通过机器学习构建最优化的模型[6-7]。本研究采用血清外泌体检测和多算法构建模型，旨在探究外泌体中miR-214-3p作为结肠癌生物标志物的潜力，并建立基于血清外泌体miR-214-3p及其他常规检验指标的结肠癌诊断模型，从而为结肠癌风险预测提供有效的补充手段。

1 资料与方法

1.1 资料

1.1.1 一般资料

本研究收集了2022年1月1日至2022年12月31日山东大学齐鲁医院普外科119例结肠癌患者（结肠癌组）和125例健康查体人群（健康对照组）的血清样本及其相关临床信息。结肠癌组纳入标准：肠镜病理报告或术后组织病理学诊断报告，均为初诊患者，未经手术、放化疗、新辅助治疗等治疗手段[8]；健康对照组纳入标准为：体检报告显示整体健康状况良好，无相关疾病史，与选取结肠癌组的年龄、性别相匹配。本研究方案通过山东大学齐鲁医院伦理委员会批准，伦理审查号：KYLL-2021（KS）-114。

1.1.2 主要仪器及试剂

HiPure Exosome RNA试剂盒购自广州美基生物科技有限公司；miRNA 1st Strand cDNA Synthesis试剂盒与Taq Pro Universal SYBR qPCR Master Mix试剂盒购自南京诺唯赞医疗科技有限公司；miR-214-3p引物由北京天根生化科技有限公司合成；PCR扩增仪购自美国BIO-RAD公司；超高速离心管与超高速离心机（Thermo MTX150）购自美国贝克曼库尔特公司；透射电子显微镜（HITACHI H-7650）购于日本日立公司；RNase-free无酶离心管与-80 ℃冰箱购自美国赛默飞世尔公司；单通道移液器购自德国eppendorf公司；电泳转模系统购自美国Bio-Rad公司；化学发光成像仪购置于美国GE公司；所用抗体均购自美国Cell Signaling Technology公司；粒径分析仪器由深圳汇芯生物医疗科技有限公司提供。

1.2 方法

1.2.1 血清外泌体提取

血清外泌体的提取采用HiPure Exosome RNA试剂盒，常规离心收集1 mL血清，并于4 ℃ 3 000 g离心15 min以去除细胞和细胞残片，转移上清液加入外泌体预沉淀溶液沉淀外泌小体，10 000×*g*离心10 min后收集外泌体。将外泌体用PBS重悬、负染、转印到载样铜网上，用透射电镜观察外泌体的形态、大小和分布。测定外泌体的蛋白浓度，将外泌体蛋白经SDS-PAGE凝胶分离，并转印到PVDF膜上，再用封闭液封闭过夜，一抗、二抗孵育，Western blotting法检测外泌体标志蛋白Alix、TSG101、CD63的表达。

1.2.2 外泌体总RNA的提取

外泌体RNA的提取采用HiPure Exosome RNA试剂盒，将收集的外泌体沉淀物储存在RNase-free离心管中，并在室温下轻轻重悬于悬浮液中，HiPure RNA Mini Column过滤重悬的外泌体，进而将含有富集RNA的HiPure RNA Mini Column转移到新的离心管中，加入无RNase水溶解RNA，并测定RNA的质量和浓度。

1.2.3 RT-qPCR检测血清外泌体miR-214-3p的表达

采用miRNA 1st Strand cDNA Synthesis试剂盒合成cDNA，根据试剂盒说明书在RNase-free离心管中制备反应体系如下：2×MiRNA RT Mix 10 μL，Hiscript miRNA Enzyme 2 μL，RNA 1 μg（根据浓度计算对应的体积），补ddH2O总体积到8 μL。反应条件：37 ℃ 60 min，85 ℃ 5 min。RT-PCR实验采用Taq Pro Universal SYBR qPCR Master Mix试剂盒，反应体系20 μL如下：2×Taq prouniversal SYBR 10 μL，universal reverse Q primer 0.2 μL，miR214-3p/U6引物0.2 μL，ddH2O 7.6 μL，CDNA 2 μL。

反应条件:95 ℃ 40 s,40 个循环扩增(95 ℃ 15 s,60 ℃ 30 s)。

1.2.4 模型建立数据选择

本研究共纳入了 31 个临床检验变量,包括血清外泌体 miR-214-3p,血常规项目(红细胞,白细胞,红细胞压积,红细胞平均宽度,血小板计数,血红蛋白,中性粒细胞比率,中性粒细胞计数,平均红细胞体积,平均血小板体积,平均血红蛋白含量,平均血红蛋白浓度,淋巴细胞比率,淋巴细胞计数,单核细胞比率和单核细胞计数),常规生化类项目(天门冬氨酸氨基转移酶,丙氨酸氨基转移酶,γ-谷丙酰基转肽酶,尿素氮,尿素肌酐比值,总胆固醇,甘油三酯,低密度脂蛋白胆固醇和高密度脂蛋白胆固醇),肿瘤标志物类(甲胎蛋白,癌胚抗原,糖类抗原 125,糖类抗原 19-9 和鳞状细胞癌相关抗原)。

1.2.5 机器学习算法

本研究采用 11 种常见的机器学习方法,包括朴素贝叶斯算法(naive_bayes),K-近邻算法(knn),逻辑回归算法(logistic_regression),随机森林算法(random_forest),决策树算法(decision_tree),人工神经网络算法(ANN),支持向量机算法(svm),集成学习梯度提升决策树(gradient_boosting),轻量级梯度提升机算法(LightGBM),自适应增强算法(AdaBoost)和极端梯度提升算法(xgboost)。本研究通过指标筛选降低模型过拟合概率,然后将 11 种分类算法的模型效果进行评价,结合 ROC 曲线和机器学习曲线选择最优算法建立模型。

1.2.6 模型建立过程

模型建立整个过程采用 python (3.10)语言进行分析处理,将研究对象按 7∶3比例随机拆分为训练集 171 例,测试集 73 例,应用不同算法对纳入的 31 个指标的重要性进行筛选、比较、综合分析,相关性大于 0.6 的指标中选择易于检测、且更便于临床常规应用的指标纳入模型,采用网格搜索方法和 10 折交叉验证方法选取每种算法的最优超参数。通过 matplotlib (3.7.1) 绘制的 ROC 曲线和 scikitplot (0.3.7)绘制的机器学习曲线对不同算法建立模型的效果进行评价,在机器学习拟合状态较好的算法中结合 ROC 曲线下面积,选取最优诊断模型。

1.3 统计学处理

采用 SPSS 23.0 软件,首先采用 Shapito-Wilk (*S-W*)方法对数据进行正态性检验,正态分布的数据以 $\bar{x}\pm s$ 表示。检验两样本对应的两总体方差是否相等:若两总体方差相等,采用 *t* 检验;若两总体方差不等,使用非参数 ANOVA 检验。对于不符合正态分布的数据,采用中位数和四分位数间距表示,两组间比较采用 Mann-Whitney *U* 检验。$P<0.05$ 为差异有统计学意义。

2 结 果

2.1 血清外泌体 miR-214-3p 在结肠癌组和健康对照组中的表达

提取的血清外泌体在透射电镜下呈直径约 30~200 nm的圆形或椭圆形小体,其边界清晰,并呈现膜包裹的囊泡样结构(图 1A),分布均匀,且颗粒直径分布图符合单峰正态分布的模式(图 1B)。此外,Western blotting 分析显示,这些颗粒表面具有外泌体标志蛋白 Alix、TSG101、CD63 的表达(图 1C)。通过提取结肠癌组和健康对照组血清外泌体,进而 RT-qPCR 扩增,结果显示外泌体 miR-214-3p 在结肠癌组的表达量显著高于健康对照组($P<0.001$,图 2)。

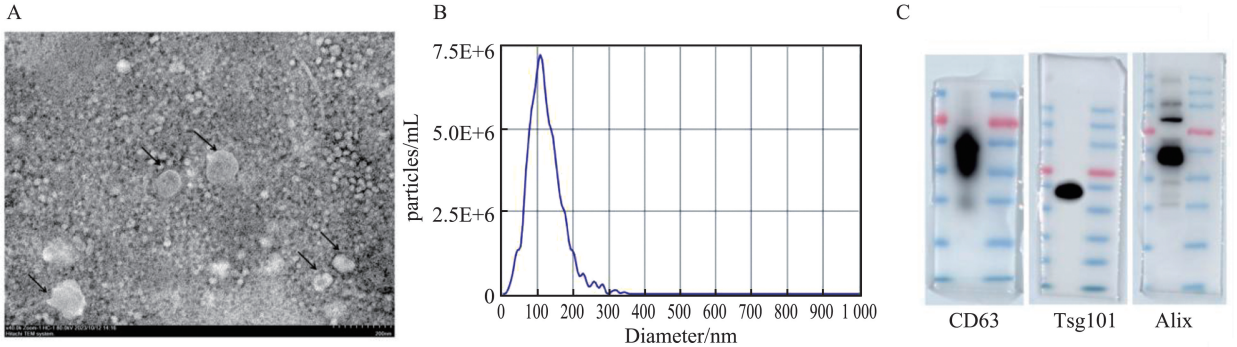


图 1 血清外泌体的鉴定

A:电镜下外泌体颗粒形态;B:颗粒直径分布图;C:Western blotting 检测结果。

Figure 1 Identification of serum exosomes

A: The shape of exosome particles under electron microscope; B: Particle diameter distribution diagram; C: Results of Western blotting.

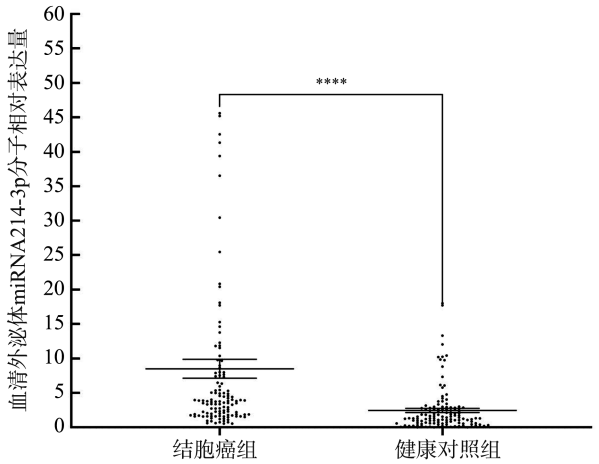


图 2 血清外泌体 miR-214-3p 分子在两组中的表达

Figure 2 The expression level of miR-214-3p in serum exosomes of colon cancer patients and healthy controls

2.2 结肠癌患者的临床信息及常规检验指标

见表 1,两组人群在年龄和性别上差异无统计学意义($P>0.05$)。同时,分析了所纳入 30 个常规检验指标在结肠癌组与健康对照组的表达差异,结果显示除平均血小板体积、低密度脂蛋白胆固醇、总胆固醇、甘油三酯、甲胎蛋白、CA-199 和鳞状细胞癌相关抗原之外,其他指标两组之间差异均具有统计学意义($P<0.05$),见表 2。

表 1 两组临床特征/n(%)

Table 1 The clinical features of 119 colon cancer patients and 125 healthy controls /n(%)

分类	结肠癌组	健康对照组
性别		
男	70(58.82)	68(54.40)
女	49(41.18)	57(45.60)
年龄/岁		
<60	55(46.22)	54(43.20)
≥60	64(53.78)	71(56.80)
肿瘤直径/cm		
≤5	67(56.30)	
>5	52(43.70)	
肿瘤位置		
左半结肠	65(54.62)	
右半结肠	54(45.38)	
TNM-pT		
T1+T2	22(18.49)	
T3+T4	97(81.51)	
TNM-pN		
N0	68(57.14)	
N1+N2	51(42.86)	
Stage 分期		
S0-SII	72(60.50)	
SIII-SIV	47(39.50)	

表 2 两组的 30 种临床检验结果分析

Table 2 Analysis of 30 clinical items of colon cancer patients and healthy controls

检验项目	结肠癌组	健康对照组	F/Z	P
血常规				
白细胞	6.90±2.67	5.76±1.38	17.949[b]	<0.001
中性粒细胞比率	64.96±11.91	56.50±8.81	39.948[b]	<0.001
中性粒细胞计数	5.73(3.93,15.75)	3.05(2.61,4.03)	−4.638[a]	<0.001
单核细胞比率	6.90(5.90, 8.10)	5.80(4.70, 6.55)	−5.645[a]	<0.001
单核细胞计数	0.43(0.33, 057)	0.32(0.26, 0.40)	−6.532[a]	<0.001
淋巴细胞比率	25.55±10.75	34.03±8.45	47.061[b]	<0.001
淋巴细胞计数	1.55(1.20, 1.84)	1.93(1.50, 2.29)	−4.654[a]	<0.001
红细胞	4.4±0.53	4.80±0.44	1.755[c]	0.186
红细胞压积	41.70(38.7,50.3)	43.60(41.50,47.30)	−8.211[a]	<0.001
红细胞平均宽度	15.00(13.20,29.90)	12.90(12.60,13.35)	−2.096[a]	0.036
血红蛋白	120.33±24.49	143.84±16.46	75.823[b]	<0.001
平均血红蛋白含量	30.35(28.30,34.60)	30.20(29.30,31.20)	−5.853[a]	<0.001
平均血红蛋白浓度	333.50(323.00,360.00)	328.00(322.00,333.00)	−2.342[a]	0.019
血小板计数	327.50(262.00,743.00)	234.00(201.50,272.50)	−2.758[a]	0.006
平均红细胞体积	90.60(86.90,105.90)	92.20(89.50,94.45)	−7.271[a]	<0.001
平均血小板体积	9.97±0.90	10.07±1.05	2.166[c]	0.142
生化指标				
γ-谷丙酰基转肽酶	16.00(11.00,24.00)	19.00(13.00,31.00)	−2.536[a]	0.011
丙氨酸氨基转移酶	11.00(7.50,15.00)	15.00(9.50,22.00)	−3.749[a]	<0.001
低密度脂蛋白胆固醇	2.96±0.69	2.91±0.79	1.132[c]	0.288

续表

检验项目	结肠癌组	健康对照组	*F*/*Z*	*P*
天门冬氨酸氨基转移酶	16.00(13.00,20.00)	18.00(15.00,21.50)	−2.616[a]	0.009
尿素	3.00(2.40,65.10)	4.80(4.10,5.45)	−8.271[a]	0.009
尿素肌酐比值	54.98±33.53	71.16±15.68	23.240[b]	<0.001
总胆固醇	4.58±0.90	4.80±0.93	0.096[c]	0.757
甘油三酯	1.13(0.88,1.44)	1.08(0.77,1.60)	−1.514[a]	0.130
高密度脂蛋白胆固醇	1.16±0.25	1.36±0.30	32.156[b]	<0.001
肿瘤标志物指标				
甲胎蛋白	2.58(1.90,3.83)	2.98(2.35,3.96)	−1.575[a]	0.115
癌胚抗原	3.29(1.90,6.55)	1.47(1.08,2.28)	−7.640[a]	<0.001
CA-125	9.96(7.09,14.20)	8.13(6.00,10.40)	−3.473[a]	0.001
CA-199	11.20(6.32,25.50)	10.62(8.22,14.34)	−0.500[a]	0.617
鳞状细胞癌相关抗原	2.05(1.62,2.71)	1.06(0.66,1.44)	−0.475[a]	0.634

[a]:非正态分布 Mann-Whitney *U* 检验;[b]:ANOVA 检验;[c]: 独立样本 *t* 检验,假定等方差。

2.3 特征性指标和不同算法筛选结果

根据随机森林算法的重要性排序,本研究选择了重要性的前 5 个指标包括尿素,外泌体 miR-214-3p,癌胚抗原,尿素肌酐比值和单核细胞计数(图 3A)。通过皮尔逊系数图分析 5 个指标的相关性分析发现,尿素和尿素肌酐比值相关性为 0.78,具有显著相关性。尿素肌酐比值需同时检测尿素和肌酐两个指标,进而计算其比值,因此构建模型时纳入尿素。最终筛选出 4 个指标:癌胚抗原,外泌体miR-214-3p,尿素和单核细胞计数(图 3B),其 AUC 值分别是 0.782、0.820、0.806和 0.742(图 4A-D)。

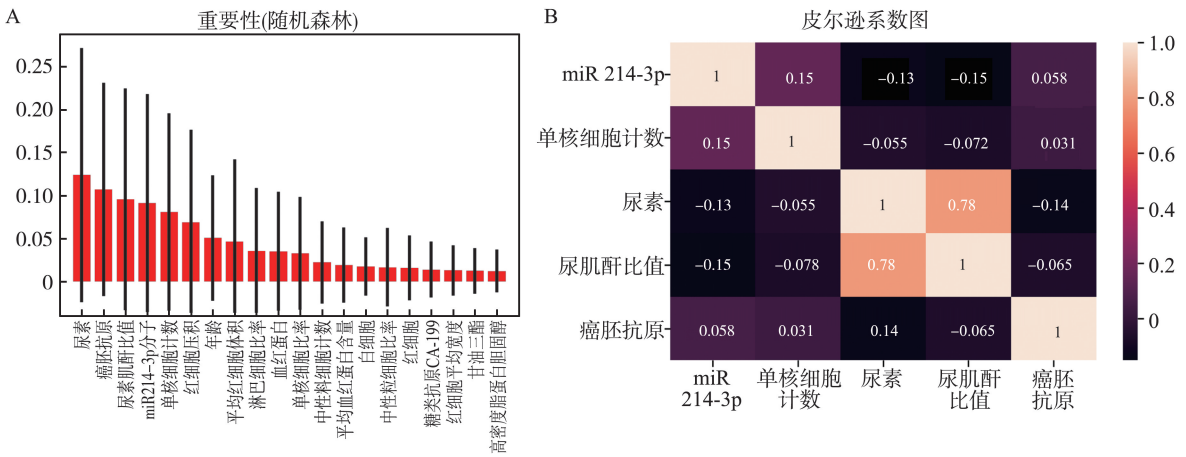


图 3 特征性指标和算法的筛选

A:随机森林算法重要性排序;B:皮尔逊相关性分析。

Figure 3 Screening of characteristic variables and algorithms

A: Random-forest algorithm importance ranking; B: Pearson correlation analysis.

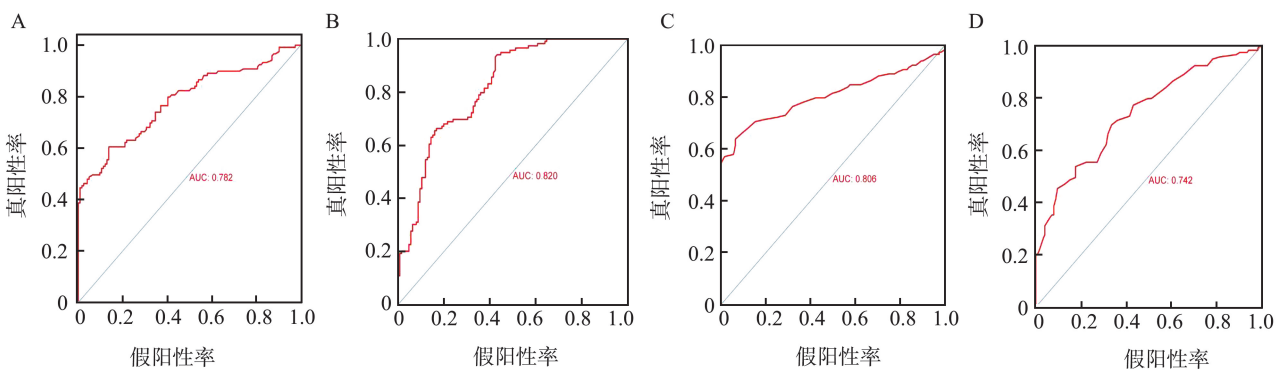


图 4 CEA(A)、外泌体 miRNA-214-3p(B)、尿素(C)和单核细胞计数(D)的诊断结肠癌的 ROC 曲线

Figure 4 ROC curves of CEA (A), serum exosomes miRNA-214-3p (B), urea (C), and monocyte count (D) in diagnosis of colon cancer

本研究将筛选出的 4 个指标采用 11 种不同算法进行建模和评价，并绘制 ROC 曲线。通过图 5A 与图 5B 的对比可以看出，减少指标后模型整体效果依然表现较好，所有算法的 AUC 值均大于 0.90。进一步绘制了这 11 种算法的学习曲线来判断模型的拟合效果，图 6A-K 显示逻辑回归算法的机器学习曲线拟合程度最好，其 AUC 为 0.93。

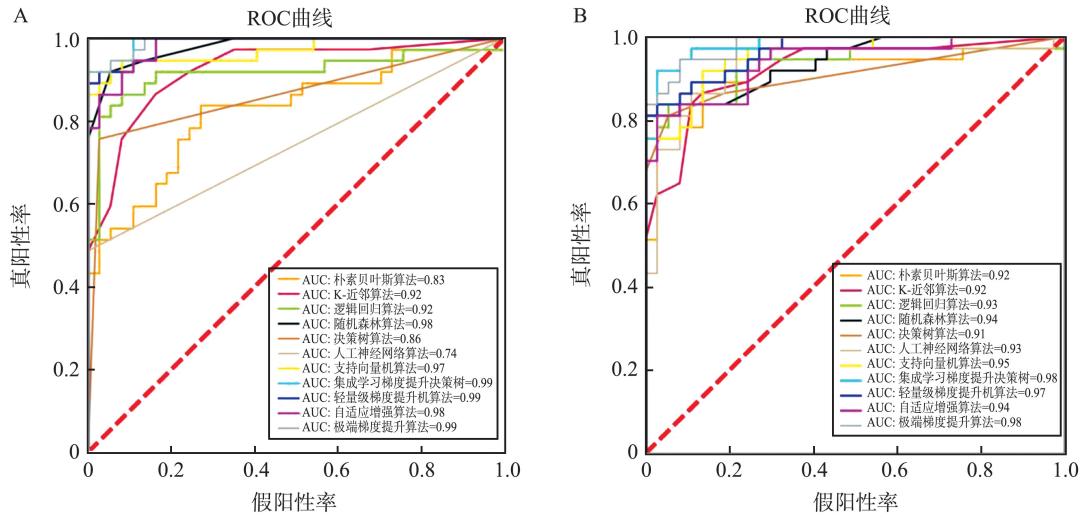
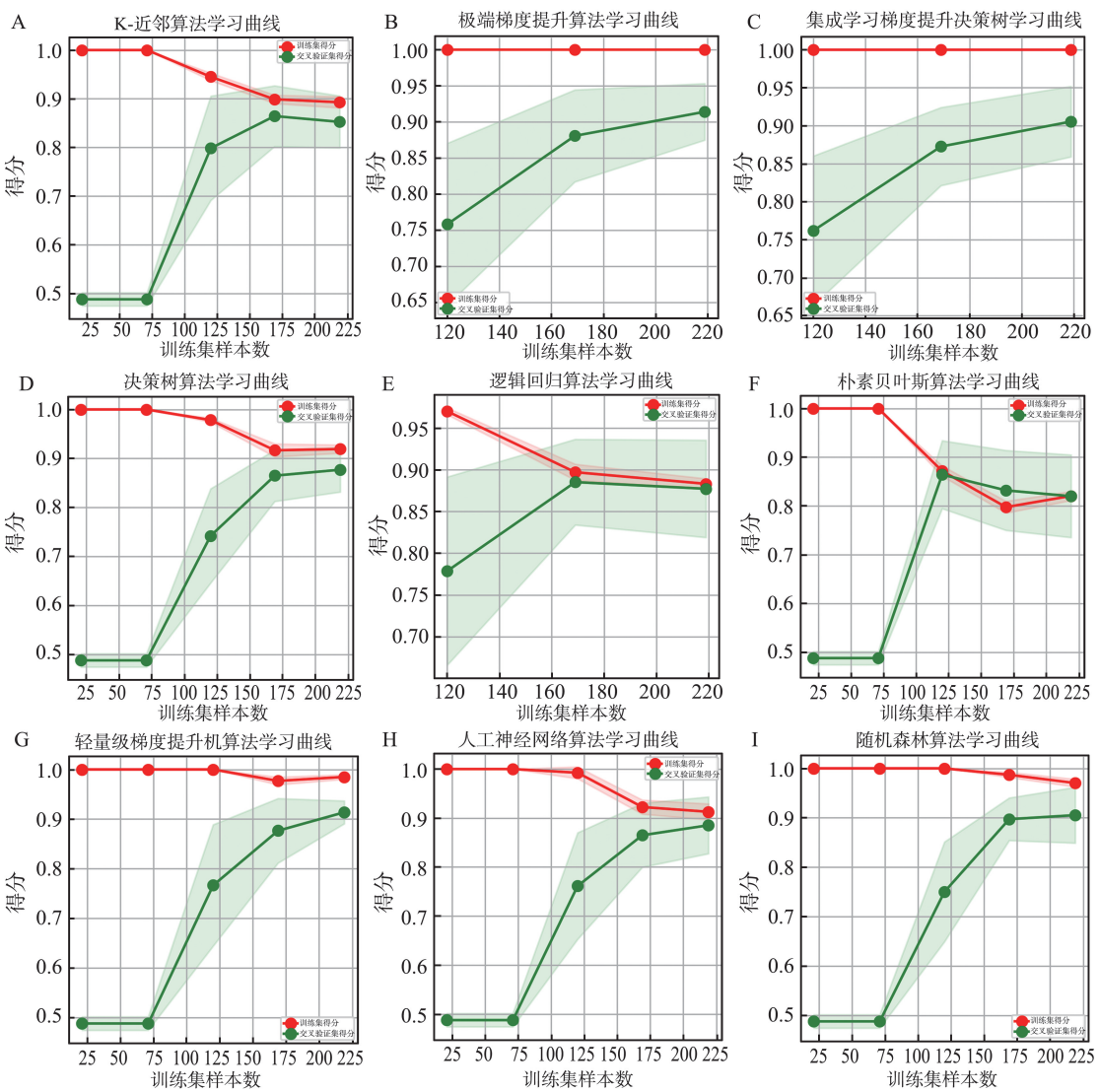


图 5 指标筛选前（A）后（B）模型评价对比

Figure 5 Comparison of model evaluation before （A） and after （B） index screening



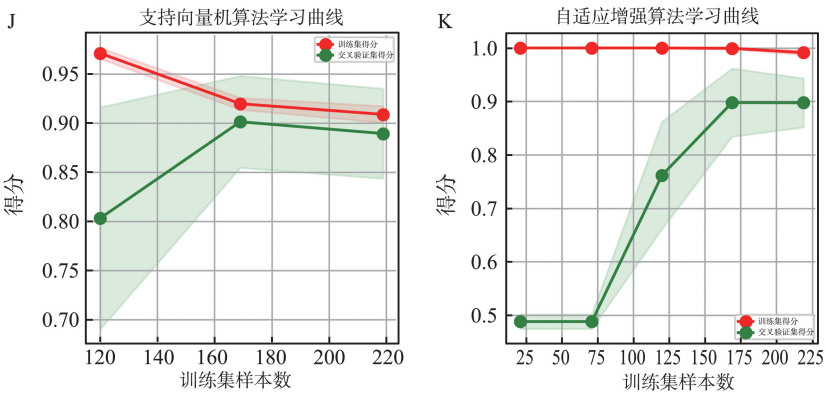


图 6 11 种算法学习曲线判断模型的拟合效果

A:K-近邻算法学习曲线;B:极端梯度提升算法学习曲线;C:集成学习梯度提升决策树学习曲线;D:决策树算法学习曲线;E:逻辑回归算法学习曲线;F:朴素贝叶斯算法学习曲线;G:轻量级梯度提升机算法学习曲线;H:人工神经网络算法学习曲线;I:随机森林算法学习曲线;J:支持向量机算法学习曲线;K:自适应增强算法学习曲线。

Figure 6 The learning curve of 11 algorithms determines the fitting effect of the model

A: Learning curve of K-nearest neighbor algorithm.; B: Extreme gradient lifting algorithm learning curve; C: Integrated learning gradient improves the decision tree learning curve; D: Decision tree algorithm learning curve; E: Logistic regression algorithm learning curve; F: Learning curve of naive Bayes algorithm; G:Lightweight gradient elevator algorithm learning curve; H: Learning curve of artificial neural network algorithm; I: Learning curve of random forest algorithm; J: Support vector machine algorithm learning curve; K: Adaptive enhancement algorithm learning curve.

2.4 逻辑回归模型的超参数

本研究筛选出的最优算法为逻辑回归算法,表 3 中列出的为逻辑回归算法的最佳超参数。

表 3 建立模型所使用的最佳超参数

Table 3 Optimal hyperparameters used to construct the model

超参数名	超参数值	超参数名	超参数值	超参数名	超参数值
C	25	intercept_scaling	1	n_jobs	None
class_weight	None	l1_ratio	None	penalty	l2
dual	False	max_iter	100	random_state	None
fit_intercept	True	multi_class	auto	solver	lbfgs
tol	0.0001	verbose	0	warm_start	False

3 讨 论

结肠癌早发现、早治疗对于改善患者预后尤为重要,早期结肠癌患者 5 年生存率可达 90%,一旦出现远处转移则仅约 15%[9-11]。然而,目前现有的结肠癌诊断方法例如癌胚抗原检测、粪便潜血试验、影像学检查和组织活检,这些方法都存在一定的局限性[12-16],因此,寻找新型的肿瘤标志物对于改善结肠癌的早期诊断至关重要。

本研究结果显示,结肠癌患者血清外泌体 miR-214-3p 的表达量显著高于健康对照组,且对于结肠癌有较好的诊断价值,其 AUC 为 0.820,优于传统结肠癌标志物 CEA 的诊断效能,可作为结肠癌诊断标志物的有效补充。肿瘤细胞来源的外泌体含有肿瘤特异性 miRNAs,血清或血浆循环外泌体中的 miRNAs 具有稳定性和可动态监测的优势,已被证实可作为潜在的肿瘤诊断、预测和治疗监测的标志物[17]。研究报道,血清外泌体 miR-377-3p 和 miR-381-3p 在结直肠癌患者中显著下调,可作为结肠癌早期诊断标志物[18];血清外泌体 miR-34a 水平在早期卵巢癌患者中显著高于晚期、有淋巴结转移或复发的患者,可作为卵巢癌的潜在生物标志物[19]。MiR-214-3p 是 miR-214 双链解旋后的 3′-端目标链,可通过靶向调节多个信号通路和基因在肿瘤的发展过程中发挥重要作用[20]。研究者发现 miR-214-3p 在肝癌细胞中充当 GPX4 的负调节剂,有助于诱导铁质沉积,从而促进肿瘤细胞的铁死亡[21];另一项研究表明,miR-214 的表达与结直肠癌肿瘤的大小、淋巴转移和预后有关,可参与结直肠癌的进展[22-23]。

本研究共纳入 119 例结肠癌患者,其中男性多于女性,年龄≥60 岁的患者较多。这与 Hossain 等[24]的研究结果一致,该研究指出男性比女性更容易罹患结肠癌,且年龄增长是危险因素之一。同时,结果显示结肠癌患者多为 T3+T4 期,表明大部分结肠癌患者难以在早期被发现,被诊断时已为中晚期。为

寻找更敏感特异性的结肠癌诊断方法,本研究纳入了31个检验变量,采用11种不同算法筛选特征性变量构建最优诊断模型,最终建立了基于尿素、癌胚抗原、单核细胞计数、外泌体miR-214-3p共4个指标的逻辑回归模型,该模型的诊断效能(AUC=0.93)显著优于目前应用广泛的标志物癌胚抗原。因此,将新型标志物与临床常规检测多检验指标相结合,采用机器学习算法构建模型对于提高结肠癌的诊断效能是非常有前景的。

本研究构建的模型中除外泌体miR-214-3p,还包括三个常规检测指标:尿素、癌胚抗原和单核细胞计数。本研究分析发现结肠癌组尿素水平降低,而癌胚抗原和单核细胞计数则显著升高。尿素是人体蛋白质代谢的主要终产物,分子量小且不与血浆蛋白结合,可自由滤过肾小球,其病理性异常可见于肾前性、肾性和肾后性因素。肾后性因素最常见肿瘤、组织坏死等伴随着蛋白代谢异常的疾病。一项研究通过对比结肠癌患者血清尿素水平和生存状况发现,血清尿素水平是结肠癌患者预后的独立因素[25-26]。血单核细胞来源于骨髓中的造血干细胞,是机体防御系统的重要组成部分,参与机体免疫反应,可识别和杀伤肿瘤细胞。当机体发生炎症或其他疾病时,单核细胞总数、百分比及细胞中的基因均可发生异常[27]。CEA是一种由胎儿胃肠道上皮组织、胰和肝细胞所合成的糖蛋白,属于非器官特异性肿瘤相关抗原,为临床常用的广谱肿瘤标志物,与多种恶性肿瘤的发病存在密切关系,在多种恶性肿瘤和良性疾病中均可升高,其诊断结肠癌的敏感性与特异性不足,一般不建议单独用于结肠癌的筛查和诊断,需要结合其他筛查手段[28]。

综上所述,血清外泌体miR-214-3p是结肠癌的潜在标志物,且采用逻辑回归算法建立的基于尿素,癌胚抗原、单核细胞计数和外泌体miR-214-3p的模型对于结肠癌的诊断具有重要价值。本研究的局限性在于纳入的检验指标有限,未能从基因组学、蛋白组学、代谢组学等多组学角度整合进行结肠癌风险预测、疾病诊断或预后判断研究。

参考文献:

[1] 闫超, 陕飞, 李子禹. 2020年中国与全球结直肠癌流行概况分析[J]. 中华肿瘤杂志, 2023, 45(3): 221-229. YAN Chao, SHAN Fei, LI Ziyu. Epidemiological analysis of colorectal cancer in China and the world in 2020. Chinese Journal of Cancer, 2019, 45(3): 221-229.

[2] Fabregas JC, Ramnaraign B, George TJ. Clinical updates for colon cancer care in 2022[J]. Clin Colorectal Cancer, 2022, 21(3): 198-203.

[3] Su Y, Tian X, Gao R, et al. Colon cancer diagnosis and staging classification based on machine learning and bioinformatics analysis[J]. Comput Biol Med, 2022, 145: 105409. doi: 10.1016/j.compbiomed.2022.105409.

[4] Tang S, Cheng J, Yao Y, et al. Combination of four serum exosomal miRNAs as novel diagnostic biomarkers for early-stage gastric cancer[J]. Front Genet, 2020, 11: 237. doi:10.3389/fgene.2020.00237.

[5] Miao C, Zhang W, Feng L, et al. Cancer-derived exosome miRNAs induce skeletal muscle wasting by Bcl-2-mediated apoptosis in colon cancer cachexia[J]. Mol Ther Nucleic Acids, 2021, 24: 923-938. doi: 10.1016/j.omtn.2021.04.015.

[6] Wei W, Li Y, Huang T. Using machine learning methods to study colorectal cancer tumor micro-environment and its biomarkers[J]. Int J Mol Sci, 2023, 24(13): 11133. doi: 10.3390/ijms241311133.

[7] Nguyen QTN, Nguyen PA, Wang CJ, et al. Machine learning approaches for predicting 5-year breast cancer survival: a multicenter study[J]. Cancer Sci, 2023, 114(10): 4063-4072.

[8] Lannagan TR, Jackstadt R, Leedham SJ, et al. Advances in colon cancer research: in vitro and animal models[J]. Curr Opin Genet Dev, 2021, 66: 50-56. doi: 10.1016/j.gde.2020.12.003.

[9] Ahmed M. Colon cancer: a clinician's perspective in 2019[J]. Gastroenterology Res, 2020, 13(1): 1-10.

[10] Khan SZ, Lengyel CG. Challenges in the management of colorectal cancer in low- and middle-income countries[J]. Cancer Treat Res Commun, 2023, 35: 100705. doi: 10.1016/j.ctarc.2023.100705.

[11] Verkuijl SJ, Jonker JE, Trzpis M, et al. Functional outcomes of surgery for colon cancer: a systematic review and meta-analysis[J]. Eur J Surg Oncol, 2021, 47(5): 960-969.

[12] 刘睿清, 卢云. 基于循证医学的早期结肠癌外科治疗进展[J]. 中华胃肠外科杂志,2022, 25(12): 1144-1149. LIU Ruiqing, LU Yun. Advances in surgical treatment of early colon cancer based on evidence-based medicine[J]. Chinese Journal of Gastrointestinal Surgery, 2002, 25(12): 1144-1149.

[13] Chan SCH, Liang JQ. Advances in tests for colorectal cancer screening and diagnosis[J]. Expert Rev Mol Diagn, 2022, 22(4): 449-460.

[14] Birgisson H, Olafsdottir EJ, Sverrisdottir A, et al. Screening for cancer of the colon and rectum a review on incidence, mortality, cost and benefit[J]. Laeknabladid, 2021, 107(9): 398-405.

[15] Jain S, Maque J, Galoosian A, et al. Optimal strategies for colorectal cancer screening[J]. Curr Treat Options Oncol, 2022, 23(4): 474-493.

[16] Shaukat A, Levin TR. Current and future colorectal cancer screening strategies[J]. Nat Rev Gastroenterol Hepatol, 2022, 19(8): 521-531.

[17] Liang G, Zhu Y, Ali DJ, et al. Engineered exosomes for targeted co-delivery of miR-21 inhibitor and chemotherapeutics to reverse drug resistance in colon cancer[J]. J Nanobiotechnology, 2020, 18(1): 10. doi: 10.1186/s12951-019-0563-2.

[18] Wang L, Song X, Yu M, et al. Serum exosomal miR-377-3p and miR-381-3p as diagnostic biomarkers in colorectal cancer[J]. Future Oncol, 2022, 18(7): 793-805.

[19] Maeda K, Sasaki H, Ueda S, et al. Serum exosomal microRNA-34a as a potential biomarker in epithelial ovarian cancer[J]. J Ovarian Res, 2020, 13(1): 47. doi: 10.1186/s13048-020-00648-1.

[20] Karimi E, Dehghani A, Azari H, et al. Molecular mechanisms of miR-214 involved in cancer and drug resistance [J]. Curr Mol Med, 2023, 23(7): 589-605.

[21] He GN, Bao NR, Wang S, et al. Ketamine induces ferroptosis of liver cancer cells by targeting lncRNA PVT1/miR-214-3p/GPX4[J]. Drug Des Devel Ther, 2021, 15: 3965-3978. doi: 10.2147/dddt.S332847.

[22] Wu Y, Xu X. Long non-coding RNA signature in colorectal cancer: research progression and clinical application [J]. Cancer Cell Int, 2023, 23(1): 28. doi: 10.1186/s12935-023-02867-0.

[23] Sukmana BI, Al-Hawary SIS, Abosaooda M, et al. A thorough and current study of miR-214-related targets in cancer[J]. Pathol Res Pract, 2023, 249: 154770. doi: 10.1016/j.prp.2023.154770.

[24] Hossain MS, Karuniawati H, Jairoun AA, et al. Colorectal cancer: a review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies[J]. Cancers (Basel), 2022, 14(7): 1732. doi: 10.3390/cancers14071732.

[25] Wang C, Sun H, Liu J. BUN level is associated with cancer prevalence[J]. Eur J Med Res, 2023, 28(1): 213. doi: 10.1186/s40001-023-01186-4.

[26] Qin WH, Yang ZS, Li M, et al. High serum levels of cholesterol increase antitumor functions of nature killer cells and reduce growth of liver tumors in mice[J]. Gastroenterology, 2020, 158(6): 1713-1727.

[27] Larionova I, Patysheva M, Iamshchikov P, et al. PFKFB3 overexpression in monocytes of patients with colon but not rectal cancer programs pro-tumor macrophages and is indicative for higher risk of tumor relapse[J]. Front Immunol, 2022, 13: 1080501. doi: 10.3389/fimmu.2022.1080501.

[28] Yu K, Qiang G, Peng S, et al. Potential diagnostic value of the hematological parameters lymphocyte-monocyte ratio and hemoglobin-platelet ratio for detecting colon cancer[J]. J Int Med Res, 2022, 50(9): 3000605221122742. doi: 10.1177/03000605221122742.

(编辑:刘霞)