技术方法 DOI:10.13406/j.cnki.cyxb.001086

基于支持向量机模型预测结直肠癌根治术预后风险

杜成凤，胡　敏，何晶晶，明　鑫，邓　丹
（重庆医科大学公共卫生与管理学院卫生统计与信息管理教研室，重庆　400016）

【摘　要】**目的**：基于支持向量机（support vector machine，SVM）预测结直肠癌患者术后风险。**方法**：回顾性分析 2011 年 1 月至 2013 年 12 月行根治性手术治疗的 199 例结直肠癌患者临床资料，按照是否无病生存（disease free survival，DFS）分为预后好、差 2 组。采用 SAS9.2 进行基本统计分析、logistic 回归分析，运用 R3.2.4 软件建立 SVM 预测模型、受试者工作曲线（receiver operating characteristic curve，ROC）及曲线下面积（area under the curve，AUC）评价模型预测效果。**结果**：差异性分析显示，总蛋白、白蛋白、阳性淋巴率、组织类型、p53、N 分期、放化疗等因素在预后组间存在统计学差异（$P<0.05$）；多因素逐步 logistic 回归结果显示，N2 分期（$OR=13.5$，95%$CI=3.85\sim47.5$）、p53 阳性（$OR=0.314$，95%$CI=0.125\sim0.790$）可能是影响预后的危险因素。线性核、多项式核、sigmoid 核及 RBF（径向基函数）核 SVM 模型的预测准确率分别为 69.70%、66.70%、69.70%、74.24%。对所有预测模型性能对比显示，SVM 模型预测性能高于 logistic 回归模型，其中 RBF-SVM 模型效果最好，预测准确率为 74.24%，特异度为 0.714，灵敏度为 0.765，AUC 值为 0.708（95%$CI=0.551\sim0.866$）。**结论**：N2 分期、p53 阳性可能是影响预后结局的危险因素，应当予以重视，减少预后风险。

【关键词】结直肠癌；风险预测；支持向量机
【中图分类号】R574.6　　【文献标志码】A　　【收稿日期】2016-07-22

Risk prediction of prognosis in patients with colorectal cancer under radical operation based on support vector machine

Du Chengfeng, Hu Min, He Jingjing, Ming Xin, Deng Dan
(*Teaching and Research Section of Health Statistics and Information Management, School of Public Health and Management, Chongqing Medical University*)

【Abstract】***Objective***: To predict the postoperative risk of patients with colorectal cancer based on support vector machine (SVM). ***Methods***: Clinical data of 199 patients with colorectal cancer were collected retrospectively from the hospital from January 2011 to December 2013, and patients were divided into two groups according to disease free survival or not. Basic statistics analysis and logistic regression were conducted by SAS9.2. Parameters optimization and establishment of SVM model were

作者介绍：杜成凤，Email：cqfddcf123@126.com，研究方向：人群健康与统计决策。
通信作者：邓　丹，Email：ddlinger@126.com。
基金项目：重庆市自然科学基金资助项目（编号：cstc2013jcyjA0068）。
优先出版：http://www.cnki.net/kcms/detail/50.1046.r.20161215.1150.038.html（2016-12-15）

[10] 巫锡炜. 中国高龄老人残障发展轨迹的类型：组基发展建模的一个应用[J]. 人口研究，2009，33(4)：54-67.
[11] Pines HA，Gorbach PM，Weiss RE，et al. Sexual risk trajectories among MSM in the United States：implications for pre-exposure prophylaxis delivery[J]. J Acquir Immune Defic Syndr，2014，65(5)：579-586.
[12] 冯国双，于石成，刘世炜. 轨迹分析模型在追踪数据分析中的应用[J]. 中国预防医学杂志，2014，15(3)：292-295.
[13] Grant RM，Lama JR，Anderson PL，et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men[J]. N Engl J Med，2010，363(27)：2587-2599.
[14] Nagin DS，Odgers CL. Group-based trajectory modeling in clinical research[J]. Annu Rev Clin Psychol，2010(6)：109-138.
[15] 付春静，钟晓妮，彭　斌，等. 男男性行为人群艾滋病暴露前预防用药临床试验队列保持力分析[J]. 重庆医科大学学报，2014，39(11)：1573-1577.
[16] 周建波，甄　森，王金塔，等. 男男性行为人群队列维持影响因素调查[J]. 中国公共卫生，2013，29(8)：1162-1165.
[17] 李洮庭，付春静，钟晓妮，等. 男男性行为者艾滋病暴露前预防用药性行为去抑制化情况[J]. 重庆医科大学学报，2015，40(6)：828-833.
[18] 闫红梅，王开利，赵东辉，等. 黑龙江省四市 MSM 人群艾滋病相关高危行为及其影响因素分析[J]. 中国艾滋病性病，2009，15(3)：288-291.
[19] 尹　宁，倪莹青，朱晓华，等. 男男性行为人群 HIV 相关行为学特征及其影响因素[J]. 环境与职业医学，2012，29(6)：397-399.
[20] 尹　宁，倪莹青，石国政，等. 上海市西部三区 MSM 人群的 HIV 相关行为学特征及影响因素[J]. 中国艾滋病性病，2012，18(11)：728-731.

（责任编辑：冉明会）

analyzed by R3.2.4. Receiver operating characteristic curve(ROC) and area under the curve(AUC) were used to evaluate their performance. ***Results***: Differences analysis showed that total protein, albumin, positive lymph node number, tissue type, N, p53, radiotherapy and chemo-therapy were significantly different between prognosis groups(P<0.05). Multiple logistic regression showed that N2(OR=13.5, 95%CI=3.85 to 47.5), p53 positive(OR=0.314, 95%CI=0.125 to 0.790) maybe the risk factors of prognosis. The prediction accuracy of linear-kernel, polynomial-kernel, sigmoid-kernel and RBF (radial basis function)-kernel was 69.70%, 66.70%, 69.70%, 74.24%, respectively. The results of performance comparison showed that SVM models performed better than logistic regression model; RBF-SVM was the best among all the models, with the prediction accuracy of 74.24%; sensitivity, specificity and AUC were 0.714, 0.765, 0.708(95%CI=0.551 to 0.866), respectively. ***Conclusion***: N2 staging and p53 positive maybe the risk factors for prognosis and more attention should be paid to decrease the risk of prognosis.

【Key words】colorectal cancer; risk prediction; support vector machine

结直肠癌(colorectal cancer, CRC)是一种常见的消化道恶性肿瘤，近年来发病率与死亡率均呈逐年上升趋势。2010 年，中国 NCCR(The National Central Cancer Registry of China)统计结果显示[1]，我国 CRC 新发病例约 27 万例，发病率为 20.10/10 万，占恶性肿瘤总体的 8.89%。2012 年，据世界卫生组织国际癌症研究中心(International Agency for Research on Cancer, IARC)资料显示[2]，全球约有 136 万 CRC 新发病例，死亡约 69 万例，CRC 已然成为威胁人类健康的重要疾病。所幸的是，随着 CRC 治疗的规范化，患者的术后无病生存率有所提高。CRC 的主要治疗手段为外科手术[3]，同时辅助放化疗治疗效果较好[4]，但术后复发转移风险仍难以避免，最常见的转移方式为肺或肝转移。一旦患者发生术后复发转移将严重影响预后[5]，故预知患者术后风险，采取适当措施尤为重要。已有研究对 CRC 术后转移[6]或患病[7]风险进行预测，但方法主要采用传统的多因素 logistic 回归和 COX 回归，以上 2 种方法对样本量要求较高，不宜分析小样本高维数据。基于统计学习理论的 SVM 学习算法[8]，适合小样本、高维数据的分类问题，是具有较好泛化能力的预测模型。该模型在图像识别[9-10]、基因蛋白表达[11-12]、疾病诊断[13]等领域应用广泛，预测效果较好，但在疾病预后风险预测中应用较少。本研究运用基于网格法参数优化的 SVM 算法及多因素 logistic 回归对 CRC 根治术后风险进行预测，采用 ROC 曲线及曲线下面积评价模型效果，以期较准确地预测患者术后风险，提高患者的生活质量，同时降低 CRC 死亡率。

1 资料与方法

1.1 资料来源

回顾性收集重庆市某三甲医院 2011 年 1 月至 2013 年 12 月入院并明确诊断的 199 例 CRC 患者资料，资料收集时间截至 2016 年 5 月 31 日。收集内容包括一般人口学特征、临床特征指标、病理学指标、免疫组化指标、血清学指标、结局指标，其中人口学指标包括性别、年龄、是否吸烟、是否饮酒；临床指标包括肿瘤部位、是否有消化道息肉、是否伴有肠梗阻、手术方式、术中放疗、是否化疗；病理学指标包括分化程度、组织分型、肿瘤大小、神经脉管是否浸润、阳性淋巴数、Dukes 分期、TNM 分期；免疫组化指标包括 CEA、CerB-2、Ki-67、GST-Π、Pgp、TOPO2、p53；血清学指标包括总蛋白、白蛋白、谷丙转氨酶、谷草转氨酶、总胆红素、直接胆红素、CA-125、CA-199、CA-242、CA-50、CA-724；结局指标包括无病生存(disease free survival, DFS)、转移、复发、死亡。

纳入标准：①经病理检查标准明确诊断为 CRC；②接受根治性手术治疗；③手术时未发现远处转移；④人口学资料及各种检查资料完整。排除标准：①非原发性 CRC；②随访时间小于 1 个月的患者；③术前已接受放化疗；④临床资料缺失。

结局判定：截至资料收集时间，患者仍处于 DFS 状态则判定为预后好；反之，若出现复发、转移、死亡或全身并发症者判定为预后差。

1.2 方法

1.2.1 基本统计分析 本研究运用 SAS 9.2 进行基本统计分析，对人口学资料及临床资料进行描述及差异性分析，以对不同预后组的年龄分布进行描述，t 检验分析组间年龄差异；以对组间的总蛋白含量、白蛋白含量、阳性淋巴数及阳性淋巴率分布进行描述，秩和检验分析组间的差异；以卡方检验分别对不同预后组的分化程度、组织类型、Duke 分期、是否化疗、是否放疗等分类资料进行差异性分析；检验水准 α=0.05。将有差异性(P<0.05)的指标纳入多因素逐步 logistic 回归模型，变量纳入标准 0.05，剔除标准 0.10。

1.2.2 支持向量机模型 支持向量机预测模型运用 R3.2.4 软件编程实现，其基本原理如下。

对于两分类问题，设训练样本集 $P=(x_i, y_i)$，$i=1, 2, \cdots\cdots n$，n 为训练样本的个数，$x_i \in R^d$ 为训练样本属性，$y_i \in \{1, -1\}$ 是输入样本 x_i 的标签。SVM 的核心问题是寻找一个最优分类平面 $g(x)=\omega x+b$，将两类样本正确分开，并使两类间边际最大，当两类样本点在二维空间中可完全分开时，分类平面

为一条直线。但在大多数情况下，两类样本数据集在二维空间有所重叠，并不能完全分开，需要利用核函数将原空间样本点映射到高维空间，使样本点在高维空间分离，得到最优分类超平面，常用的核函数有线性核(linear)、多项式核(polynomial)、高斯核(RBF)和 Sigmoid 核函数。

采用网格搜索算法优化核函数参数 c 和 γ，然后结合 10 折交叉验证(10-fold cross validation)得到交叉验证准确率，选择准确率最高的 c 和 γ 作为核函数参数。若多组 c 和 γ 均可得到最高准确率时，选择参数 c 值最小的那组作为最佳的参数组合，避免因 c 过高导致过拟合现象。利用最优参数 c 和 γ 建立预测模型，以回代样本分类准确率、预测样本分类准确率、特异度、灵敏度、ROC 曲线下面积(AUC)评价模型预测性能。其中 ROC 曲线可直观显示出模型区分两类目标的能力大小，是较常用的一类评价指标。AUC 越接近 1，表明分类性能越好。

2 结 果

2.1 患者基本情况

研究共纳入 199 名 CRC 患者，其中男性 126 例(63.32%)，女性 73 例(36.68%)，平均年龄(60.69 ± 10.00)岁，确诊时年龄最大者 89 岁，最小者 28 岁。直肠癌患者 113 名，占56.78%，结肠癌患者 86 名，占 43.22%，组织类型以管状腺癌为主(占 86.43%)，其余为黏液腺癌及少数印戒细胞癌，患者 Dukes 分期主要为 B 期(45.23%)与 C 期(42.71%)，少数 A 期(12.06%)。截至观察结束，共 61 人出现观察结局，其中 36 人出现肝转移，9 人出现复发，6 人腹腔转移，4 人死亡，3 人肺转移，2 人广泛转移，1 人脑转移，其余 138 人处于无病生存状态。

2.2 预后与临床特征差异性分析

研究纳入指标共 35 个，其中肿瘤标志抗原 CA-199、CA-125、CA-242、CA-50、CA-724 及 CEA 含量以 P25(第 25 百分位数)、P75(第 75 百分位数)为界划分为低、中、高 3 组；阳性淋巴率为阳性淋巴数与送检总淋巴数之比；其余因素以原始值纳入分析。差异性分析可初步筛选可能的预后危险因素。分析结果显示，年龄、阳性淋巴数、阳性淋巴率、总蛋白含量、白蛋白含量、组织类型、N 分期、p53、Dukes 分期、术中放疗、术后化疗在预后分组间存在统计学差异($P<0.05$)，可以认为以上 11 个因素可能是预后的危险因素。详细结果见表1。

表 1 预后与患者资料的差异性分析

Tab.1 Differential analysis between prognosis and clinical characteristics

变量名	预后		$t/z/\chi^2$ 值	P 值
	好(n=138)	差(n=61)		
年龄(岁)	59.52 ± 12.71	63.34 ± 11.67	2.01	0.046
白蛋白(g/L)	36.45 ± 7.90	34.10 ± 6.90	-2.02	0.022
阳性淋巴数(个)	0.00 ± 1.00	1.00 ± 4.00	-3.45	0.001
阳性淋巴率(%)	0.00 ± 0.16	0.11 ± 0.46	-3.31	0.001
组织分型(例)				
管状腺癌	121	51		
黏液腺癌	10	10	6.7	0.015
其他腺癌	7	0		
n(例)				
0	87	27		
1	42	17	17.5	0
2	9	17		
p53(例)				
阳性	66	40	5.35	0.021
阴性	72	21		
Duke分期(例)				
A	19	5		
B	68	22	6.21	0.045
C	51	34		
术中放射粒子植入量(n)				
0 粒	38	8		
10 粒	4	0	8.81	0.012
20 粒	96	53		
是否术后化疗(例)				
否	19	18	6.92	0.001
是	119	43		

2.3 预测模型建立

2.3.1 训练及测试样本的确定 将 199 例样本按 2:1 的比例随机分配到训练集和测试集，其中训练集共 133 例（预后差 43 例，占 32.33%），测试集共 66 例（其中预后差 18 例，占 27.27%）。

2.3.2 建立多因素 logistic 回归模型 根据差异性分析结果，共 11 个因素纳入多因素逐步回归预测模型，变量纳入标准设置为 0.05，剔除标准为 0.10。分类变量具体赋值情况见表 2，其中对组织分型、Duke 分期、术中放疗及 N 分期等多分类变量均设置了亚变量。多因素逐步 logistic 回归结果显示，p53 为阳性（OR=0.314，95%CI=0.125~0.790，P=0.014）的CRC 患者预后越好，N2 期患者（OR=13.5，95%CI=3.85~47.5）比N0 期患者更容易倾向预后差的结局；提示 N2 分期和 p53 可能是预后结局的危险因素。具体结果见表 3。

2.3.3 SVM 模型 运用网格搜索算法优化 SVM 模型的参数 c、γ，将参数的优化范围以一定的步长划分为多个网格，分别将每个网格内的参数带入模型进行训练，经过多次尝试和对比，得到交叉验证准确率最好的参数值为 c=0.16，γ=0.062 5；然后以 133 例训练集为基础，分别使用 Linear 核、Polynomial 核、Sigmoid 核及 RBF 核函数建立预测模型，用训练得到的模型代入训练样本数据可得到回代准确率。

2.4 模型预测效果比较

利用 66 例测试集样本对建立的 logistic 回归模型、Linear-SVM、Polynomial-SVM、Sigmoid-SVM 及 RBF-SVM 5 种预测模型分别进行测试，预测效果用准确率、特异度、灵敏度及 ROC 曲线下面积来评价。对比结果见表 4：Linear-SVM、Polynomial- SVM、Sigmoid-SVM 及 RBF-SVM 拟合的预测模型预测准确率均高于 logistic 回归模型（66.67%）。另外，除了 RBF 核函数以外，其余 4 种模型都出现了一定程度的过拟合现象；而 RBF-SVM 拟合的模型不仅未出现过拟合现象，并且准确率高于其他 4 种模型（74.24%），模型预测效果最佳。

根据图 1~图 5 结果可看出，Polynomial-SVM、Linear-SVM、Sigmoid-SVM 3 种模型的灵敏度和特异度相同（Sp=0.776，Se=0.647）；RBF-SVM 模型的灵敏度高于前面 3 种模型，特异度稍低；logistic 回归模型灵敏度显著低于前 4 种模型（Se=0.333）。模型 AUC 值由高到低进行排名为 RBF-SVM、

表 2 分类变量赋值表

Tab.2 Assignment table of categorical variable

变量名		赋值
自变量	组织分型	(100)=管状腺癌；(010)=黏液腺癌；(001)=其他
	N	(100)=2；(010)=1；(001)=0
	p53	0=阴性；1=阳性
	Duke 分期	(100)=C 期；(010)=B 期；(001)=A 期
	术后是否化疗	0=否；1=是
	术中放疗	(100)=20 粒；(010)=10 粒；(001)=未放疗
因变量	预后分组	0=好；1=差

表 3 多因素逐步 logistic 回归结果

Tab.3 Result of multiple stepwise logistic regression

参数		β	β标准误	$Wald\chi^2$	P	OR	OR 95%CI	
							下限	上限
截距		−0.356	0.438	0.664	0.415			
N	以 N0 为对照							
	N2	2.600	0.642	16.500	0.000	12.500	3.850	47.500
	N1	0.268	0.463	0.334	0.563	0.074	0.021	0.260
p53		−1.260	0.471	6.050	0.014	0.314	0.125	0.790

表 4 模型训练结果

Tab.4 Training results of models

核函数	交叉验证准确率（%）		特异度 Sp	灵敏度 Se	AUC	AUC 95%CI
	训练集	测试集				
Linear-SVM	75.94	69.70	0.776	0.647	0.670	(0.505，0.835)
Polynomial-SVM	78.20	66.70	0.776	0.647	0.672	(0.508，0.837)
Sigmoid-SVM	75.94	69.70	0.776	0.647	0.670	(0.505，0.835)
RBF-SVM	70.68	74.24	0.714	0.765	0.708	(0.551，0.866)
logistic	74.44	66.67	0.765	0.333	0.549	(0.412，0.686)

Polynomial-SVM、Linear-SVM、Sigmoid-SVM、logistic 回归模型。综上所述，SVM 拟合的 4 种模型预测性能均较 logistic 回归模型好，在 SVM 模型中 RBF-SVM 拟合的预测模型效果最佳。

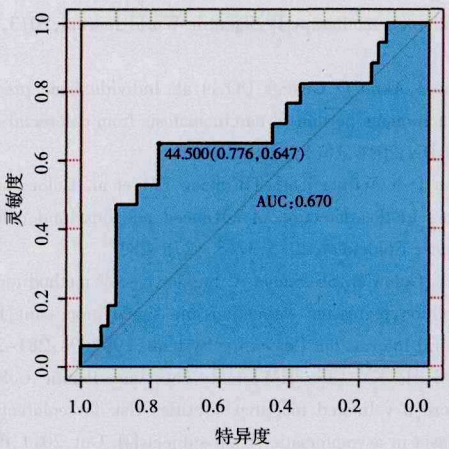


图 1　线性核 ROC 曲线图

Fig.1　ROC curve of linear kernel function

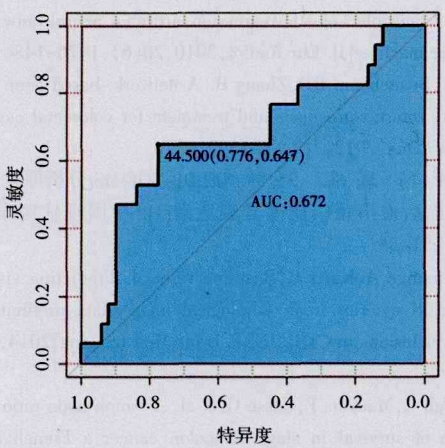


图 2　多项式核 ROC 曲线图

Fig.2　ROC curve of Polynomial kernel function

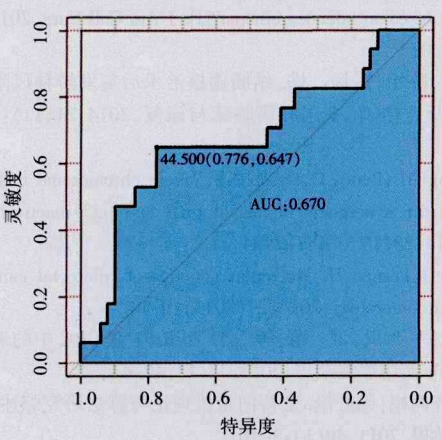


图 3　sigmoid 核 ROC 曲线图

Fig.3　ROC curve of sigmoid kernel function

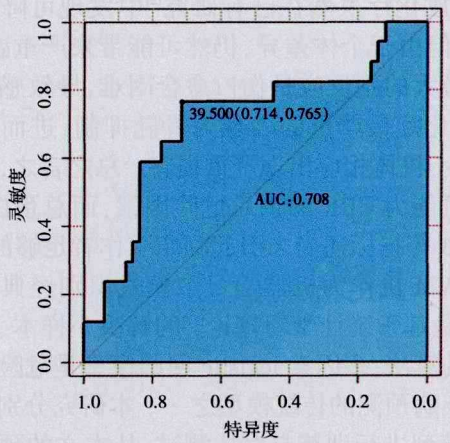


图 4　RBF 核 ROC 曲线图

Fig.4　ROC curve of RBF kernel function

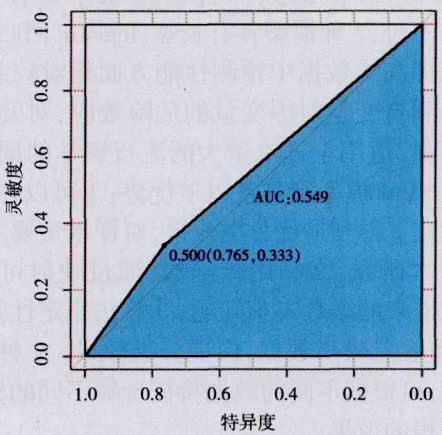


图 5　Logistic 回归 ROC 曲线图

Fig.5　ROC curve of logistic regression

3　讨　论

本研究通过建立预测模型，发现可能影响 CRC 术后结局的因素共 11 个，其中淋巴转移率[14-15]、年龄[15]、分化程度[15]、Duke 分期[16]是在以往研究中已得到证实的预后危险因素。研究还发现，p53 蛋白的过表达也可能是术后的危险因素，p53 是人类 17 号染色体上的抑癌基因，其蛋白的过表达是染色体不稳定的标志[17]，而染色体不稳定是 CRC 形成的途径之一[18]。虽然国内已有研究表明 p53[19]在结直肠腺癌和腺瘤中的表达存在差异（P=0.019），推测 p53 蛋白表达或许与预后有关，但尚缺乏前瞻性临床试验证据。另外，本研究还发现总蛋白含量、白蛋白含量可能与预后有关，此 2 种蛋白不仅可反映肝脏代谢及合成功能，同时由于进展期的恶性肿瘤过多地消耗蛋白而出现降低，故可间接反映肿瘤生长的状态。除此之外，结果中显示放化疗可能是预后结局的影响

因素，放化疗虽然在已有研究[4]中发现可降低肿瘤复发，但由于个体差异，仍然可能带来严重副反应。放疗最大的副反应是伤口愈合困难，导致感染或复发；而化疗最严重的表现为骨髓抑制，进而全血细胞减少，机体相应出现严重损害。总之言之，以上因素均可能为 CRC 的预后危险因素，而总蛋白、白蛋白、p53 等指标还尚未引起临床工作者足够的重视。

SVM 机作为机器学习中模式识别经典算法之一[20]，是基于统计学习理论[17]的解决小样本、高维度数据的算法；多因素 logistic 回归模型是危险因素筛选及疾病预测的传统模型之一。本研究分别将 2 种经典模型进行训练与对比测试，从本文的预测结果可看出，SVM 预测性能均较 logistic 回归模型好，其中 RBF-SVM 核函数拟合模型效果最佳，AUC=0.708。不过，2 种模型各有长短，logistic 回归模型虽然在处理高维数据中预测性能方面相对较差，但结果中得到自变量对因变量的危险效应，对变量的解释性更强，适用于样本量大的流行病学的调查类数据。而 SVM 模型则具备以下优势：①可以在有限的样本信息下得到最优分类效果，对样本量要求较低，更适合本研究；②利用核函数，通过映射可避免高维数据带来的维数灾难问题；③算法稳定性强、鲁棒性好，可容忍噪声数据，保证分类效果。2 种模型各有千秋，可根据不同的数据特征选择不同的模型，达到较理想的效果。

综上所述，年龄、总蛋白、白蛋白、淋巴转移率、放化疗、p53、分化程度以及 Duke 分期等因素可能是影响 CRC 患者术后的危险因素，其中部分指标需给予足够的重视；RBF-SVM 拟合的预测模型涉及较多指标，较全面地包含影响预后的临床因素，对预测患者根治术风险有一定的参考价值。虽然本研究建立的模型达到了一定的预测效果，发现了可能影响患者预后的因素，但仍存在一些问题。由于观察性研究本身存在一定的局限性，另外，参数优化方法的单一导致部分模型仍出现了过拟合现象。项目组将进一步结合遗传算法及粒子群算法优化参数，以期获得预测性能更好的模型，改善 CRC 患者预后。

参 考 文 献

[1] Zheng ZX, Zheng RS, Zhang SW, et al. Colorectal cancer incidence and mortality in China, 2010[J]. Asian Pac J Cancer Prev, 2014, 15(19): 8455-8460.

[2] Stewart BW, Wild CP. World cancer report 2014[R]. Switzerland: the International Agency for Research on Cancer, 2014.

[3] NCNC. NCCN guidelines rectal cancer and colon cancer[DB/CD]. https://www.nccn.org/ professionals/physician_gls/f_guidelines.asp. 2015.

[4] Malik AI, Badar F, Syed AA, et al. Surgically treated rectal cancer patients d outcomes at a tertiary care cancer hospital in Pakistan[J]. Asian J Surg, 2015, 38(1): 13-20.

[5] Yonemura Y, Canbay E, Ishibashi H. Prognostic factors of peritoneal metastases from colorectal cancer following cytoreductive surgery and perioperative chemotherapy[J]. Scientiic World Journal, 2013, 2013: 978394.

[6] Segelman J, Akret O, Gustafs UO, et al. Individualized prediction of risk of metachronous peritoneal carcinomatosis from colorectal cancer [J]. Colorectal Dis, 2014, 16(5): 359-367.

[7] Stegeman I, de Wijkerslooth TR, Stoop EM, et al. Colorectal cancer risk factors in the detection of advanced adenoma and colorectal cancer[J]. Cancer Epidemiol, 2013, 37(3): 278-283.

[8] Vapnik V, Golowich SE, Smola A. Support vector method for function approximation, regression estimation, and signal processing[J]. Advances in Neural Information Processing Systems, 1997(9): 281-287.

[9] Yeoh KG, Ho KY, Chiu HM, et al. The Asia-Pacific Colorectal Screening score: a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects[J]. Gut, 2011, 60(9): 1236-1241.

[10] Lee HJ, Hwang SI, Han SM, et al. Image-based clinical decision support for transrectal ultrasound in the diagnosis of prostate cancer: comparison of multiple Logistic regression, artificial neural network, and support vector machine[J]. Eur Radiol, 2010, 20(6): 1476-1484.

[11] Shi M, Beauchamp RD, Zhang B. A network-based gene expression signature informs prognosis and treatment for colorectal cancer patients[J]. PLoS One, 2012, 7(7): e41292.

[12] 齐义军，马 瑾，张 天，等. SELDI-TOF/MS 分析原发性肝癌 TACE 治疗前后血清蛋白质差异表达谱[J]. 中国现代医学杂志，2014，24(5)：51-55.

[13] Depeursinge A, Kurtz C, Beaulieu C, et al. Predicting visual semantic descriptive terms from radiological image data: preliminary results with liver lesions in CT[J]. IEEE Trans Med Imaging, 2014, 33(8): 1669-1676.

[14] Sabbagh C, Mauvais F, Cosse C, et al. A lymph node ratio of 10% is predictive of survival in stage III colon cancer: a French regional study[J]. Int Surg, 2014, 99(4): 344-353.

[15] Gleisner AL, Mogal H, Dodson R, et al. Nodal status, number of lymph nodes examined, and lymph node ratio: what defines prognosis after resection of colon adenocarcinoma?[J]. J Am Coll Surg, 2013, 217(6): 1090-1100.

[16] 周雄坤，陈祖林，杨 桦. 结肠癌根治术后复发转移风险模型的初步建立及评估[J]. 中国肿瘤临床与康复，2014，21(11)：1301-1303.

[17] Manning AL, Benes C, Dyson NJ. Whole chromosome instability resulting from the synergistic effects of pRB and p53 inactivation[J]. Oncogene, 2014, 33(19): 2487-2494.

[18] Bogaert J, Prenen H. Molecular genetics of colorectal cancer[J]. Annals of Gastroenterology, 2014, 27(2014): 9-14.

[19] 叶 斌，王西川，雷 蕾，等. p53 和 Ki67 在 CRC 中的表达及其临床意义[J]. 实用癌症杂志，2013，28(1)：42-44.

[20] 丁世飞，齐丙娟，谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报，2011，40(1)：3-10.

（责任编辑：冉明会）