

• 短篇论著 •

常规体检项目构建的支持向量机模型在宫颈癌筛查中的价值研究*

余鑫,晏沙沙,张曦,朱锐,何亮[△]
(云南省肿瘤医院检验科,云南昆明 650000)

摘要:目的 探索基于常规体检项目(血常规、尿常规及生化常规)建立的支持向量机模型在宫颈癌筛查中的意义。方法 通过文献筛查筛选出可能与肿瘤相关的血常规指标 6 项:白细胞计数(WBC)、血小板计数、淋巴细胞计数、中性粒细胞计数(NEUT)、中性粒细胞/淋巴细胞比值(NLR)、血小板/淋巴细胞比值(PLR);通过小样本初筛筛选出可能与宫颈癌相关的生化常规及尿常规指标 7 项:尿比重、pH、总蛋白(TP)、尿酸、钙(Ca)、铁、镁;比较 40 例宫颈癌患者(宫颈癌组 B)和 40 例健康体检者(对照组 B)上述 13 项指标水平,筛选出差异有统计学意义的指标,用于建立宫颈癌筛查的支持向量机模型。结果 两组 WBC、NEUT、NLR、PLR、TP、Ca 水平比较,差异均有统计学意义($P < 0.05$)。用上述 6 项指标构建支持向量机模型,其中筛查模型诊断宫颈癌的灵敏度为 86.11%,特异度为 88.46%,诊断符合率为 87.10%;验证模型诊断宫颈癌的灵敏度为 66.67%,特异度为 100.00%,诊断符合率为 83.33%。受试者工作特征曲线分析结果显示,构建的支持向量机模型诊断宫颈癌的曲线下面积为 0.933。结论 运用常规体检项目构建的支持向量机模型具有较高的宫颈癌诊断价值,能有效提高宫颈癌早期诊断的准确性。

关键词:宫颈癌; 支持向量机模型; 血常规; 生化常规; 尿常规**DOI:**10.3969/j.issn.1673-4130.2020.17.022**中图法分类号:**R711.74**文章编号:**1673-4130(2020)17-2147-04**文献标识码:**B

宫颈癌是发生于宫颈上皮的与人乳头状瘤病毒(HPV)感染相关的一种恶性肿瘤^[1]。国际癌症研究所(IARC)数据显示,全球每年宫颈癌新发病例约为 53.0 万,占有女性肿瘤的 12.0%,死亡病例达 27.5 万^[2]。而我国每年约有 13.0 万宫颈癌新发病例,5.3 万女性死于该病,宫颈癌已经成为严重威胁女性身体健康的恶性肿瘤之一^[3]。宫颈癌及其癌前病变是可预防的,但由于发展中国家缺乏经济、技术基础,宫颈癌筛查措施不完善,患者常失去早期治疗的机会^[4-5]。

目前,宫颈癌早期筛查主要依靠液基薄层细胞学技术和 HPV 检测等^[6],但上述方法在发展中国家尚未普及,且因为价格较高等原因导致主动筛查的人数较少,迫切需要寻找更为经济、便利、高效的筛查方法。血常规、尿常规、生化常规检测属于常规体检项目,具有价格便宜、标本留取方便、检测方法简单等优点,适用于多种疾病的筛查。支持向量机模型属于神经网络分析模型,是一种模拟人类大脑信息处理方式的简化模型。支持向量机算法基于经验风险最小化原则,同时强调置信范围最小,采用结构风险最小化原理,整合泛化能力与算法复杂度之间的矛盾。因此,支持向量机算法在小样本研究领域的表现优于传统模式的识别方法^[7]。目前,已有较多研究将神经网络分析模型用于生物医学^[8-9]。本研究通过分析宫颈

癌患者和健康人群之间血常规、尿常规及生化常规检测水平间的差异,应用支持向量机算法建立宫颈癌筛查模型,用于宫颈癌和健康人群的鉴别,以期精确识别宫颈癌高危人群提供可靠途径。

1 资料与方法

1.1 一般资料 选取本院妇科 2016 年 1 月 1 日至 2018 年 12 月 31 日收治的宫颈癌患者 50 例作为宫颈癌组。纳入标准:(1)病理学或细胞学检查初次确诊为宫颈癌;(2)20~70 岁女性;(3)入院后接受相关影像学检查,如腹部彩超、胸部 CT、头部 MRI、全身骨扫描或 PET-CT,明确肿瘤 TNM 分期;(4)临床资料完整。排除标准:(1)合并其他类型肿瘤;(2)既往进行过放疗或化疗;(3)合并基础疾病,如糖尿病、高血压、冠心病等。另选取同期年龄为 20~70 岁的健康体检女性 50 例作为对照组。宫颈癌组和对照组中分别随机选取 10 例研究对象用于小样本初筛,两组分别剩余的 40 例研究对象用于模型构建。

1.2 仪器与试剂 生化常规检测试剂和仪器由罗氏诊断产品(上海)有限公司提供;血常规检测试剂和仪器由日本希森美康株式会社提供。

1.3 方法

1.3.1 文献筛查 从文献中筛选出与肿瘤相关的血常规指标^[10-13]。

* 基金项目:国家癌症中心肿瘤科研专项课题(NCC2017A33);云南省科技厅科技计划项目[2018EF001(-248)]。

[△] 通信作者, E-mail: fishxin725@163.com。

1.3.2 小样本初筛 收集 10 例健康体检者(对照组 A)和 10 例宫颈癌患者(宫颈癌组 A)的生化常规及尿常规指标,进行比较,筛选出差异有统计学意义的指标,即为宫颈癌相关生化常规及尿常规指标。指标纳入标准:完整的生化常规指标包括肝功能、肾功能、蛋白、电解质、血糖、血脂、心肌酶等相关指标。指标排除标准:非常规体检中心开展项目,如胱抑素 C 等;人工操作项目,如尿镜检细菌等。根据指标纳入及排除标准,最终纳入的生化常规初筛指标:钾、钠、氯、钙(Ca)、镁(Mg)、铁(Fe)、总蛋白(TP)、清蛋白、总胆红素、直接胆红素、丙氨酸氨基转移酶(ALT)、天门冬氨酸氨基转移酶(AST)、AST/ALT、尿素、尿酸(UA)、肌酐、血葡萄糖(GLU)、三酰甘油、胆固醇、低密度脂蛋白、高密度脂蛋白、肌酸激酶、肌酸激酶同工酶。纳入的尿常规初筛指标:pH、尿比重(SG)、尿胆原、隐血、尿蛋白、尿 GLU、尿胆红素、尿酮体、尿红细胞。

1.3.3 筛选模型构建指标 比较剩余 40 例宫颈癌患者(宫颈癌组 B)和 40 例健康体检者(对照组 B)小样本初筛筛选出的生化常规与尿常规指标,以及文献筛查所筛选出的肿瘤相关血常规指标,将差异有统计学意义的指标用于模型构建。

1.4 统计学处理 采用 SPSS25.0 软件进行数据分

析。正态性验证采用 Kolmogorov-Smirnov 检验;正态分布的计量资料以 $\bar{x} \pm s$ 表示,组间比较采用 t 检验;不符合正态分布的计量资料以 $M(P_{25}, P_{75})$ 表示,组间比较采用秩和检验;采用受试者工作特征曲线(ROC 曲线)评价支持向量机模型诊断宫颈癌的效能。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 宫颈癌相关指标小样本初筛 从相关文献中筛选出可能与肿瘤相关的血常规指标,包括白细胞计数(WBC)、血小板计数(PLT)、中性粒细胞/淋巴细胞比值(NLR)、血小板/淋巴细胞比值(PLR)、淋巴细胞计数(LY)、中性粒细胞计数(NEUT)。通过小样本初筛,筛选出宫颈癌组 A 与对照组 A 间差异有统计学意义的指标,即为宫颈癌相关的生化常规及尿常规指标,包括 SG、pH、TP、UA、Ca、Fe、Mg。见表 1。

2.2 筛选模型构建指标 将文献筛查筛选出的可能与肿瘤相关的血常规指标及表 1 中宫颈癌相关生化常规及尿常规指标在宫颈癌组 B 和对照组 B 间进行比较,结果显示,WBC、NEUT、NLR、PLR、TP、Ca 水平在两组间比较,差异均有统计学意义($P < 0.05$),可纳入模型构建。见表 2。

表 1 两组间差异有统计学意义的生化常规及尿常规指标比较结果

组别	n	SG	pH	TP	UA	Ca	Fe	Mg
		$(\bar{x} \pm s)$	$[M(P_{25}, P_{75})]$	$(\bar{x} \pm s, g/L)$	$(\bar{x} \pm s, \mu mol/L)$	$(\bar{x} \pm s, mmol/L)$	$(\bar{x} \pm s, mmol/L)$	$(\bar{x} \pm s, mmol/L)$
宫颈癌组 A	10	1.015±0.006	5.75(5.00,6.13)	77.10±4.82	301.00±79.40	2.45±0.09	26.23±8.25	0.929±0.075
对照组 A	10	1.026±0.011	6.25(6.00,7.50)	66.80±10.97	229.90±55.19	2.27±0.17	10.34±4.75	0.853±0.051
t/Z		-2.811	-2.058	2.718	2.325	2.694	4.501	2.563
P		0.012	0.043	0.018	0.033	0.016	<0.001	0.021

表 2 两组模型构建指标比较结果

组别	n	WBC	PLT	LY	NEUT
		$[M(P_{25}, P_{75}), \times 10^9/L]$	$[M(P_{25}, P_{75}), \times 10^9/L]$	$[M(P_{25}, P_{75}), \times 10^9/L]$	$[M(P_{25}, P_{75}), \times 10^9/L]$
对照组 B	40	6.86(5.27,8.83)	263.00(207.75,312.25)	1.62(1.26,2.21)	4.73(3.53,5.81)
宫颈癌组 B	40	5.76(4.46,6.37)	239.00(176.00,290.75)	1.87(1.62,2.36)	3.17(2.46,3.65)
Z		1.901	0.894	1.342	2.348
P		0.001	0.400	0.055	<0.001

组别	n	NLR	PLR	Fe	Mg	Ca
		$[M(P_{25}, P_{75})]$	$[M(P_{25}, P_{75})]$	$[M(P_{25}, P_{75}), mmol/L]$	$(\bar{x} \pm s, mmol/L)$	$[M(P_{25}, P_{75}), mmol/L]$
对照组 B	40	2.80(2.13,4.21)	164.65(126.00,219.69)	12.05(5.45,19.60)	0.87±0.07	2.29(2.12,2.40)
宫颈癌组 B	40	1.44(1.26,2.02)	116.62(87.90,155.76)	18.30(13.15,22.25)	0.89±0.07	2.35(2.31,2.46)
t/Z		-2.565	-1.606	1.901	1.342	0.224
P		0.053	0.112	0.001	0.055	1.000

组别	n	SG	pH	TP	UA
		$[M(P_{25}, P_{75})]$	$[M(P_{25}, P_{75})]$	$[M(P_{25}, P_{75}), g/L]$	$[M(P_{25}, P_{75}), \mu mol/L]$
对照组 B	40	1.02(1.01,1.03)	6.00(5.50,6.50)	74.50(64.50,79.00)	256.50(214.00,312.75)
宫颈癌组 B	40	1.01(1.01,1.02)	6.00(5.50,6.50)	76.00(74.00,80.00)	277.50(248.25,350.00)
t/Z		1.565	1.342	2.683	1.453
P		0.015	0.055	<0.001	0.029

2.3 支持向量机模型的构建及验证

2.3.1 样本处理 从宫颈癌组 B 与对照组 B 的 80 例样本中随机选出 62 例用于构建筛查模型, 剩余 18 例用于构建验证模型。

2.3.2 支持向量机模型的构建 模型输入层以 6 个特征变量(WBC、NEUT、NLR、PLR、TP、Ca)作为协变量, 单元数为 6; 隐藏层为随机拟合的 9 个变量, 目的是将非线性不可分的输入信号设法变换成线性可分信号; 输出层节点则是利用线性函数对输入信号进行简单的线性加权和变换, 用于鉴别宫颈癌患者和健康人群。

2.3.3 支持向量机模型评价结果 筛查模型诊断宫颈癌的灵敏度为 86.11%, 特异度为 88.46%, 诊断符合率为 87.10%; 验证模型诊断宫颈癌的灵敏度为 66.67%, 特异度为 100.00%, 诊断符合率为 83.33%, 见表 3。ROC 曲线分析结果显示, 构建的支持向量机模型诊断宫颈癌的曲线下面积(AUC)为 0.933, 见图 1。

表 3 支持向量机模型评价结果

模型		病理确诊(n)		诊断符合率 (%)
		宫颈癌	健康人群	
筛查模型	宫颈癌	31	3	87.10
	健康人群	5	23	
验证模型	宫颈癌	6	0	83.33
	健康人群	3	9	

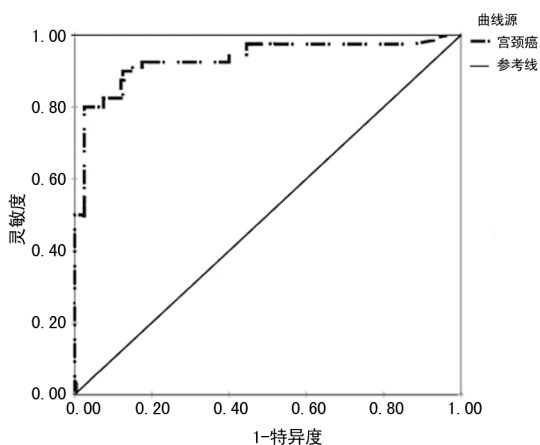


图 1 支持向量机模型诊断宫颈癌的 ROC 曲线

3 讨论

宫颈癌是严重危害女性健康的恶性肿瘤之一, 发病率占全球女性恶性肿瘤的第 2 位, 仅次于乳腺癌^[14]。目前, 宫颈癌筛查最佳方案包括, (1) 优选方案: 液基薄层细胞学技术和 HPV 联合筛查, 此筛查方法灵敏度、特异度高, 但价格昂贵、技术要求高, 仅适用于经济发达地区; (2) 一般方案: 巴氏涂片与 HPV 联合筛查, 此方法快速、便捷, 但价格也较高, 适用于

中等发达地区; (3) 基本方案: 肉眼观察法, 此方法假阳性率高, 但价格低廉、易于操作, 适用于医疗资源落后地区^[15]。研究表明, 85% 的宫颈癌新发病例和 90% 的死亡病例发生在医疗资源落后地区^[16]。宫颈癌前病变缺乏有效治疗时易发展成为宫颈癌, 早期发现与治疗是宫颈癌防治的关键。

虽然常规体检项目与液基薄层细胞学技术、HPV、宫颈癌相关肿瘤标志物检测相比缺乏特异性, 但如果能在患者未表现出明显症状时通过分析其在常规体检项目中的相关指标变化来预测宫颈癌的发生, 对于提高早期筛查率具有重要意义。故本研究筛选出血常规、尿常规、生化常规中对宫颈癌具有一定诊断意义的指标, 并通过支持向量机算法, 建立了宫颈癌筛查模型, 以期宫颈癌的筛查提供一定参考。本研究结果显示, 根据文献筛查、小样本初筛、模型构建指标筛选所选出的血常规、尿常规、生化常规指标所构建的支持向量机模型诊断宫颈癌的 AUC 为 0.933, 诊断效能高, 能够较为准确地诊断宫颈癌, 值得在临床推广应用。

此外, 神经网络分析模型还可用于其他肿瘤的模型预测。通过建立各类肿瘤的神经网络分析模型, 整合大数据库, 能帮助临床医师更有针对性地开展检查项目, 提高诊断效率; 同时, 也可为患者减少不必要的检查, 做到医疗资源的合理使用。

4 结论

运用常规体检项目构建的支持向量机模型具有较高的宫颈癌诊断价值, 能有效提高宫颈癌早期诊断的准确性。

参考文献

- [1] 祁德波, 张丽冉, 赵东曼, 等. 高危型 HPV 病毒载量与不同宫颈病变的相关性分析[J]. 检验医学与临床, 2019, 16(15): 2127-2130.
- [2] 韩耀风, 高宪利, 周莹, 等. 基于递归系统模型的宫颈癌患者住院费用影响因素分析[J]. 中国卫生统计, 2017, 34(2): 253-256.
- [3] 刘萍. 中国大陆 13 年宫颈癌临床流行病学大数据评价[J]. 中国实用妇科与产科杂志, 2018, 34(1): 41-45.
- [4] 王利丽, 马丽丽, 尚丽新. 宫颈癌及癌前病变筛查策略的研究进展[J]. 医学综述, 2016, 22(10): 1907-1910.
- [5] 裴蕴峰, 经先振, 周娟, 等. 医院就诊人群 HPV 感染的流行病学特征及其与宫颈癌/癌前病变的关系[J]. 华中科技大学学报(医学版), 2018, 47(3): 349-353.
- [6] BOLGOVA L S. About the monograph "Screening and differentiated cytomorphicologic diagnostic of cervix cancer" by V. I. Novik[J]. Klin Lab Diagn, 2014, 18(2): 63-64.
- [7] 吴疆. 基于支持向量机的红外谱图识别[D]. 成都: 四川大学, 2005.

- [8] 刘奕. 基于机器学习的癌症诊断方法研究[D]. 武汉: 湖北工业大学, 2017.
- [9] 郭利斌. 基于支持向量机算法的癌症组织拉曼光谱数据分析[D]. 福州: 福建师范大学, 2015.
- [10] 陈欣, 张国楠. 早期宫颈癌患者手术后外周血淋巴细胞计数变化的临床意义[J]. 中华妇产科杂志, 2014, 49(3): 208-212.
- [11] 唐红, 吴伟莉, 金凤, 等. 外周血中性粒细胞及血小板与淋巴细胞的比值与局部晚期鼻咽癌预后及疗效的相关性[J]. 肿瘤防治研究, 2019, 46(1): 32-36.
- [12] AMRIT P S R, KAUR M, ZONUNSANGA B, et al. Pre-operative peripheral blood count in breast carcinoma: predictor of prognosis or a routine test[J]. Int J Breast Canc-

er, 2015, 2015: 964392.

- [13] 黄利军, 方佳峰, 吴珏堃, 等. 术前血清肿瘤标志物联合血常规指标对结肠癌患者预后评估的价值[J]. 中华胃肠外科杂志, 2018, 21(12): 1421-1426.
- [14] 王燕. 新疆喀什贫困地区维吾尔族妇女子宫颈癌及癌前病变的危险因素分析[D]. 石河子: 石河子大学, 2018.
- [15] 许驰, 何玉. 宫颈癌筛查方法的现状及进展[J]. 蚌埠医学院学报, 2018, 43(11): 1538-1541.
- [16] 周晖, 王东雁, 罗铭, 等. 《FIGO 2018 妇癌报告》——子宫颈癌指南解读[J]. 中国实用妇科与产科杂志, 2019, 35(1): 95-103.

(收稿日期: 2020-02-12 修回日期: 2020-06-11)

• 短篇论著 •

1 例引起突发性耳聋的猪链球菌的鉴定及分子流行病学分析*

郑礼杰¹, 李欣霖², 张文², 陈建国¹, 谈忠鸣³, 许金凤^{1△}

(1. 江苏大学附属人民医院, 江苏镇江 212002; 2. 江苏大学医学院, 江苏镇江 212013;

3. 江苏省疾病预防控制中心, 江苏南京 210009; 4. 镇江市疾病预防控制中心, 江苏镇江 212000)

摘要:目的 分离鉴定 1 例突发性耳聋患者血液和脑脊液分离的猪链球菌并进行分子流行病学分析。方法 采用 VITEK-2 Compact 微生物系统进行初步鉴定和药敏试验, PCR 扩增细菌的 16S rRNA 基因、7 个管家基因及 4 种毒力基因。通过实时荧光定量 PCR 进行血清学分型, 多位点序列分型 (MLST) 进行序列分型 (STs)。结果 传统细菌鉴定方法和 16S rRNA 基因测定方法均鉴定为猪链球菌, 实时荧光定量 PCR 鉴定为猪链球菌 2 型, MLST 分型为 ST7, 毒力因子溶血素基因、细胞外蛋白因子编码基因、荚膜多糖编码基因均为阳性。结论 引起患者突发性耳聋的细菌为我国常见的猪链球菌 2 型, 属于 ST7 型强毒株。

关键词:猪链球菌; 突发性耳聋; 多位点序列分型; 毒力因子

DOI:10.3969/j.issn.1673-4130.2020.17.023

中图法分类号:R446.5

文章编号:1673-4130(2020)17-2150-05

文献标识码:B

猪链球菌是一种兼性厌氧的革兰染色阳性细菌, 寄生在猪、牛、驴等哺乳动物、鸟类以及人类中, 可导致脑膜炎、败血症、心内膜炎、肺炎、关节炎等疾病, 是一种重要的人畜共患病原菌^[1-3]。近二十年, 由猪链球菌引起的人类感染显著增加, 被认为是新出现的动物传染性疾病^[4]。2002—2013 年全球共有 34 个国家 1 642 例猪链球菌病的报道, 90.2% 发生在亚洲, 8.5% 发生在欧洲^[5], 西方国家病例主要与职业接触有关, 而相当一部分亚洲病例则与食用生猪肉制品有关^[6]。根据荚膜多糖 (CPS) 抗原性的不同, 猪链球菌被分为 35 种血清型 (1—34 型以及 1/2 型)^[5]。目前, 对猪致病性较强的常见血清型有 1 型、2 型、1/2 型、7 型和 9 型等^[7]。根据 7 个管家基因 (aroA、cpn60、dpr、gki、muts、recA、thrA) 核苷酸序列的不同, 应用

多位点序列分型 (MLST) 技术将猪链球菌分为不同的序列类型 (STs)^[8], 目前已知的 STs 已有 700 多种。毒力相关基因纤溶酶释放相关蛋白编码基因 (mrp)、溶血素基因 (sly)、细胞外蛋白因子编码基因 (epf)、荚膜多糖编码基因 (cps) 等在猪链球菌致病过程中发挥重要作用^[9-10]。本研究对分离自江苏大学附属人民医院 (后简称“本院”) 的 1 例突发性耳聋患者血液和脑脊液中的猪链球菌进行病原学鉴定和药敏试验, 并进行分子流行病学分析, 现报道如下。

1 资料与方法

1.1 病例介绍 患者, 男, 59 岁, 因“突发性耳聋”入院就诊。入院时头部胀痛、头晕、听力下降、言语不清、双下肢乏力, 伴有恶心、呕吐。辅助检查: 白细胞 $5.7 \times 10^9/L$, 中性粒细胞计数 $5.3 \times 10^9/L$, 中性粒细

* 基金项目: 江苏省社会发展重点研发项目 (BE2017693); 镇江市社会发展项目 (SH2019057); 镇江市第一人民医院院级课题 (Y2019010)。

△ 通信作者, E-mail: 642095775@qq.com。