文章编号:1000－1638(2025)01-0069-08 DOI:10.13484/j.nmgdxxbzk.20250108

基于307例宫颈癌患者临床特征的生存状态预测分析*

孙鹏哲,许春洁,闫祖威,冯永娥
(内蒙古农业大学理学院,呼和浩特 010018)

摘要:通过探索宫颈癌患者的临床特征与生存状态之间的关系,构建准确可靠的生存状态预测模型,可以为宫颈癌的预防、诊断和治疗提供新的思路和方法。本文基于TCGA数据库307例宫颈癌患者的临床随访数据,运用Spearman相关性分析探讨宫颈癌临床特征的相关性,首次运用因子分析探究了宫颈癌临床特征的基本结构,并提出基于因子分析的二元Logistic回归生存状态预测模型。模型预测总精度为84.2%,召回率为84.2%,精确度为81.0%,模型的AUC值为0.861,该模型在分类任务上表现良好,具有较高的预测能力和可靠性。

关键词:宫颈癌;生存状态;因子分析;Logistic回归
中图分类号:Q61 **文献标志码**:A

宫颈癌是常见女性生殖系统恶性肿瘤之一。调查研究发现,尽管通过接种HPV疫苗、早期筛查、手术治疗等方法使得宫颈癌患者病死率大大降低,但是其发病率和死亡率均居女性恶性肿瘤的高位,且近年来有年轻化趋势,严重威胁女性的生命健康和生活质量[1]。在许多卫生系统薄弱的国家和地区,宫颈癌仍保持着较高的发病率和死亡率[2-3]。全球每年约有57万例宫颈癌新发病例,大约有25万女性死于宫颈癌,其中80%的死亡病例分布于发展中国家[4-5]。世界卫生组织(WHO)在2020年发布了《加速消除宫颈癌全球战略》,得到了全球多个国家的响应和支持。我国印发了《加速消除宫颈癌行动计划(2023—2030年)》,加快推进我国宫颈癌消除进程,保护和促进广大妇女健康[6-7]。近年来,在筛查、预防和治疗宫颈癌方面已经取得巨大进展,但防治之路仍任重而道远。辛雪焕[8]认为有大量宫颈癌患者发现时已处于晚期,或者部分患者出现复发和转移,治疗效果差,且晚期宫颈癌患者的5年生存率仍然较低,生存状态预测是目前研究的热门课题。

改善患者生存结局仍然是临床诊疗实践的核心目标。经过长期临床观察和临床研究发现,即便临床分期相同的患者,生存预后亦存在显著个体化差异[9-10]。所以迫切需要建立个体化生存预测模型,以便尽早评估患者预后不良的概率,为临床医生制定相应诊疗方案提供参考。张美琴等[11]发现35岁以下年轻妇女宫颈癌的临床分期、淋巴管是否累及和宫颈肌层浸润深度是影响预后的独立因素。Rajeevan等[12]对HPV-16阳性的妇女探究CIN3的危险因素,发现高龄、抽烟和高病毒载量等是危险因素。彭俊等[13]通过对90例年轻宫颈癌患者临床特征及术后复发和预后相关因素分析,发

* **收稿日期**:2024-07-09; **修回日期**:2024-11-11
基金项目:国家自然科学基金项目(62262050,32160258);内蒙古自治区教育科学研究"十四五"规划课题项目(NGJGH2024053)
作者简介:孙鹏哲(1979－),女,内蒙古呼和浩特人,副教授,2020级博士研究生。主要研究方向为生物信息学、应用统计学。E-mail:1325095864@qq.com
通信作者:冯永娥(1977－),女,内蒙古呼和浩特人,教授,博士。主要研究方向为理论生物学、生物信息学。E-mail:yefeng@imau.edu.cn

现临床分期、宫颈浸润深度和盆腔淋巴结转移是影响研究组和对照组 5 年生存率的共同因素。此外，Wu 等[14]研究发现，绝经、家族肿瘤史和阴道清洁度都对 HPV 感染有影响。Wright 等[15]发现随着肿瘤分级的升高，患者的 5 年生存率逐步下降。Jeong 等[16]提出淋巴结转移也是宫颈癌最重要的预后因素，最近被纳入国际妇产科联合会(HGO)分期系统。Kahn 等[17]通过随机数生成器从潜在的对照组患者中随机选择，得出宫颈癌更有可能是绝经前、当前或以前的吸烟者。Macios 等[18]针对细胞学正常但 HPV 感染情况未知的妇女探讨宫颈癌的危险因素，得到阴道微生物感染、抽烟和宫颈形状与宫颈癌有关的结论。孙君华等[19]研究表明，中青年宫颈癌患者死亡与 FIGO 分期、分化程度、组织学分级、淋巴结转移情况等因素有关。

宫颈癌临床数据的复杂性主要体现在相关因素的交错和非线性关系。但是，几乎所有临床肿瘤研究首选的生存分析方法，仍然选择使用以传统线性假设为基础的 Cox 比例风险回归模型(CPH)。临床实践中潜在风险因素与死亡率的非线性相关现象更为普遍，有待发掘和研究。因子分析将具有错综复杂关系的对象综合为少数几个潜因子，寻求基本结构，简化观测系统。Logistic 回归常用于数据挖掘，疾病自动诊断等领域。本文通过探究宫颈癌的临床特征，了解其特征差异，基于多种数据分析方法，建立新型生存状态预测模型，结果模型预测总精度为 84.2%，召回率为 84.2%，精确度为 81.0%，模型的 AUC 值为 0.861，可见该模型具有较高的预测能力，对下一步提高宫颈癌的有效治疗和治疗方案的选择有积极意义。

1 研究数据与方法

1.1 数据来源

本文基于 TCGA 数据库 307 例宫颈癌患者的临床随访数据，结合相关文献研究结果，选取包括初次病理诊断年龄(X_{Age})、女性更年期状态(X_{Mpe})、怀孕总数(X_{Pgt})、生育活产儿的成功妊娠量(X_{Fty})、癌细胞是否已经扩散到远离原发灶的其他部位(X_{M})、淋巴结受累情况(X_{N})、原发肿瘤的分类(X_{T})、生长部位级别(X_{Gth})、最后一次随访患者的肿瘤状态(X_{Tmr})、病人所处癌症的具体临床阶段(X_{Stg})、吸烟程度(X_{Smk})、身高(X_{Hgt})、体重(X_{Wgt})、生存状态(X_{Svl})共 14 个指标。

1.2 研究方法

本文首先通过 Spearman 相关分析探究宫颈癌的临床指标的相关性，通过因子分析提取临床特征的潜因子，最后运用基于因子分析的二元 Logistic 回归模型对患者的生存状态进行预测分析。

1.3 评价指标

本文运用以下指标来评估模型的性能，即精确度(Precision，Pre)、召回率(Recall，Rec)、预测总精度(Accuracy，Acc)，F_1 分数(F_1)，受试者工作特征(Receiver operating characteristic，ROC)曲线下面积(Area under the curve，AUC)，分别定义为

$$Pre=\frac{TP}{TP+FP},Rec=\frac{TP}{TP+FN},Acc=\frac{TP+TN}{TP+FN+TN+FP},F_1=2\times\frac{Pre\times Rec}{Pre+Rec},$$

其中，TP 表示被模型预测正确的正样本数量，TN 表示被模型预测正确的负样本数量，FP 表示被模型预测错误的正样本数量，FN 表示被模型预测错误的负样本数量，N 表示样本总数，AUC 值是 ROC 曲线下的面积[20]。

2 结果与分析

2.1 临床数据指标的 Spearman 相关性分析

基于 TCGA 数据库宫颈癌患者的临床随访数据，对宫颈癌临床 13 个指标进行 Spearman 相关性分析，结果如表 1 所示。

由表 1 可知，病人所处癌症的具体临床阶段(Stg)与癌细胞是否已经扩散到远离原发灶的其他部位(M)、淋巴结受累情况(N)、原发肿瘤的分类(T)呈现相关关系，初次病理诊断年龄(Age)与女性更

年期状态(Mpe)呈现显著相关关系,最后一次随访患者的肿瘤状态(Tmr)与癌细胞是否已经扩散到远离原发灶的其他部位(M)呈现相关关系,淋巴结受累情况(N)与原发肿瘤的分类(T)呈现相关关系,怀孕总数(Pgt)与生育活产儿的成功妊娠量(Fty)呈现显著相关关系。

表 1　临床指标 Spearman 相关性分析

Table 1　Spearman correlation analysis of clinical indicators

指标	Stg	Age	Mpe	Pgt	Fty	Tmr	N	M	T	Gth	Smk	Hgt	Wgt
Stg	1.000	0.193**	0.134*	0.089	0.188**	0.182**	0.384**	0.434**	0.816**	0.006	0.031	−0.125*	−0.163**
Age	0.193**	1.000	0.761**	0.220**	0.275**	0.109	0.114*	−0.085	0.245**	0.058	−0.045	−0.262**	−0.064
Mpe	0.134*	0.761**	1.000	0.173*	0.203**	0.057	0.019	−0.135	0.145*	0.030	−0.074	−0.138*	−0.022
Pgt	0.089	0.220**	0.173*	1.000	0.802**	0.079	0.139*	0.015	0.043	−0.114	−0.086	−0.157*	0.121
Fty	0.188**	0.275**	0.203**	0.802**	1.000	0.089	0.248**	0.055	0.161	−0.045	−0.082	−0.200**	0.019
Tmr	0.182**	0.109	0.057	0.079	0.089	1.000	0.264**	0.337**	0.202**	0.019	0.009	−0.051	−0.032
N	0.384**	0.114*	0.019	0.139*	0.248**	0.264**	1.000	0.341**	0.458**	−0.007	0.024	−0.217**	−0.144
M	0.434**	−0.085	−0.135	0.015	0.055	0.337**	0.341**	1.000	0.213*	−0.045	0.086	0.076	0.077
T	0.816**	0.245**	0.145*	0.043	0.161	0.202**	0.458**	0.213*	1.000	−0.037	−0.006	−0.117	−0.258**
Gth	0.006	0.058	0.030	−0.114	−0.045	0.019	−0.007	−0.045	−0.037	1.000	−0.005	−0.050	−0.020
Smk	0.031	−0.045	−0.074	−0.086	−0.082	0.009	0.024	0.086	−0.006	−0.005	1.000	0.159*	0.084
Hgt	−0.125*	−0.262**	−0.138*	−0.157*	−0.200**	−0.051	−0.217**	0.076	−0.117	−0.050	0.159*	1.000	0.262**
Wgt	−0.163**	−0.064	−0.022	0.121	0.019	−0.032	−0.144	0.077	−0.258**	−0.020	0.084	0.262**	1.000

注:* 和 ** 分别表示在 $P<0.05$ 和 $P<0.01$ 水平下差异显著。

2.2　基于因子分析的二元 Logistic 回归模型对生存状态的预测

2.2.1　宫颈癌临床指标的因子分析

通过因子分析提取潜因子,可以寻求宫颈癌临床特征的基本结构,简化观测系统。同时,通过计算各因子的权重,可以评估不同因子对整体数据的重要性,且因子分析的结果能够为后续研究提供方向。结合临床指标与临床阶段的相关性研究结果,选取初次病理诊断年龄(X_{Age})、女性更年期状态(X_{Mpe})、怀孕总数(X_{Pgt})、生育活产儿的成功妊娠量(X_{Fty})、癌细胞是否已经扩散到远离原发灶的其他部位(X_{M})、淋巴结受累情况(X_{N})、原发肿瘤的分类(X_{T})、生长部位级别(X_{Gth})、最后一次随访患者的肿瘤状态(X_{Tmr})、病人所处癌症的具体临床阶段(X_{Stg})、吸烟程度(X_{Smk})11 个指标进行因子分析。

(1) 因子分析适合性检验

KMO 和巴特利特检验结果显示,KMO 的值为 0.615,巴特利特球形度检验的显著性 P 值为 0.000,因此拒绝原假设,变量间具有相关性,适合进行因子分析。

(2) 提取公因子

公因子的方差解释率如表 2 所示。由表 2 可知,公因子按照方差百分比进行了排序,方差百分比越大排序越靠前,说明该公因子对原始数据解释能力越强。因子分析是对原始数据进行降维,提取对原始数据解释能力强的公因子。前 5 个公因子的累积方差解释率为 72.636%,说明这 5 个潜因子较好地包含了原始数据的大部分信息,在保证公因子累积解释方差贡献率较大的情况下,并结合实际分析选择前 5 个公因子构建因子分析模型。

(3) 因子旋转

利用方差最大化正交旋转对所得因子载荷矩阵进行旋转,使得提取出的因子具有更好的解释性,

旋转后因子载荷系数如表 3 所示,分别将 1—5 公因子命名为癌症临床阶段因子、更年期状态因子、妊娠因子、肿瘤转移扩散因子和生长部位级别因子。

表 2 因子总方差

Table 2 Factor total variances

公因子	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差/%	累积/%	总计	方差/%	累积/%	总计	方差/%	累积/%
1	2.689	24.445	24.445	2.689	24.445	24.445	1.949	17.717	17.717
2	1.766	16.052	40.496	1.766	16.052	40.496	1.756	15.966	33.683
3	1.437	13.066	53.563	1.437	13.066	53.563	1.712	15.561	49.244
4	1.113	10.118	63.681	1.113	10.118	63.681	1.420	12.913	62.157
5	0.985	8.955	72.636	0.985	8.955	72.636	1.153	10.478	72.636
6	0.915	8.321	80.957						
7	0.906	8.241	89.197						
8	0.463	4.211	93.409						
9	0.288	2.616	96.025						
10	0.255	2.317	98.342						
11	0.182	1.658	100.000						

表 3 旋转后因子载荷系数

Table 3 Factor load coefficients after rotation

指标	因子 1	因子 2	因子 3	因子 4	因子 5
X_{Age}	0.080	0.917	0.137	0.031	0.111
X_{Pgt}	0.026	0.076	0.898	−0.037	−0.012
X_{Mpe}	0.060	0.933	0.056	−0.052	−0.037
X_{Fty}	0.089	0.117	0.894	0.180	0.006
X_{Gth}	0.062	0.033	0.116	−0.136	0.708
X_{Smk}	−0.038	0.034	−0.126	0.263	0.664
X_{Tmr}	0.058	−0.041	−0.038	0.624	0.187
X_{M}	0.059	0.028	0.154	0.687	−0.383
X_{N}	0.483	0.011	0.155	0.639	0.111
X_{T}	0.907	0.139	0.034	0.139	−0.030
X_{Stg}	0.929	0.013	0.053	0.086	0.035

(4) 因子得分

依据因子得分系数结果(表 4),计算因子得分,其中 ZX_{Age}、ZX_{Mpe}、ZX_{Pgt}、ZX_{Fty}、ZX_{M}、ZX_{N}、ZX_{T}、ZX_{Gth}、ZX_{Tmr}、ZX_{Stg}、ZX_{Smk}均为标准化后的变量。因子得分结果如下:

$$F1=-0.046\times ZX_{\mathrm{Age}}-0.025\times ZX_{\mathrm{Pgt}}-0.031\times ZX_{\mathrm{Mpe}}-0.042\times ZX_{\mathrm{Fty}}+0.039\times ZX_{\mathrm{Gth}}-0.104\times ZX_{\mathrm{Smk}}-0.115\times ZX_{\mathrm{Tmr}}-0.126\times ZX_{\mathrm{M}}+0.127\times ZX_{\mathrm{N}}+0.504\times ZX_{\mathrm{T}}+0.537\times ZX_{\mathrm{Stg}}$$

$$F2=0.533\times ZX_{Age}-0.069\times ZX_{Pgt}+0.557\times ZX_{Mpe}-0.039\times ZX_{Fty}-0.04\times ZX_{Gth}+0.02\times ZX_{Smk}-0.004\times ZX_{Tmr}+0.045\times ZX_{M}-0.027\times ZX_{N}+0.013\times ZX_{T}-0.074\times ZX_{Stg}$$

$$F3=-0.03\times ZX_{Age}+0.563\times ZX_{Pgt}-0.08\times ZX_{Mpe}+0.532\times ZX_{Fty}+0.106\times ZX_{Gth}-0.084\times ZX_{Smk}-0.077\times ZX_{Tmr}+0.009\times ZX_{M}+0.015\times ZX_{N}-0.054\times ZX_{T}-0.02\times ZX_{Stg}$$

$$F4=0.037\times ZX_{Age}-0.122\times ZX_{Pgt}-0.017\times ZX_{Mpe}+0.044\times ZX_{Fty}-0.135\times ZX_{Gth}+0.24\times ZX_{Smk}+0.501\times ZX_{Tmr}+0.535\times ZX_{M}+0.394\times ZX_{N}-0.1\times ZX_{T}-0.156\times ZX_{Stg}$$

$$F5=0.055\times ZX_{Age}+0.018\times ZX_{Pgt}-0.078\times ZX_{Mpe}+0.03\times ZX_{Fty}+0.62\times ZX_{Gth}+0.576\times ZX_{Smk}+0.163\times ZX_{Tmr}-0.332\times ZX_{M}+0.088\times ZX_{N}-0.06\times ZX_{T}+0.003\times ZX_{Stg}$$

表 4　因子得分系数表

Table 4　Factor score coefficients

指标	因子				
	$F1$	$F2$	$F3$	$F4$	$F5$
ZX_{Age}	−0.046	0.533	−0.030	0.037	0.055
ZX_{Pgt}	−0.025	−0.069	0.563	−0.122	0.018
ZX_{Mpe}	−0.031	0.557	−0.080	−0.017	−0.078
ZX_{Fty}	−0.042	−0.039	0.532	0.044	0.030
ZX_{Gth}	0.039	−0.040	0.106	−0.135	0.620
ZX_{Smk}	−0.104	0.020	−0.084	0.240	0.576
ZX_{Tmr}	−0.115	−0.004	−0.077	0.501	0.163
ZX_{M}	−0.126	0.045	0.009	0.535	−0.332
ZX_{N}	0.127	−0.027	0.015	0.394	0.088
ZX_{T}	0.504	0.013	−0.054	−0.100	−0.060
ZX_{Stg}	0.537	−0.074	−0.020	−0.156	0.003

2.2.2　基于因子分析的二元 Logistic 回归模型

依据因子分析中提取的 5 个潜因子，即癌症临床阶段因子、更年期状态因子、妊娠因子、肿瘤转移扩散因子、生长部位级别因子，利用其因子得分指标 $F1$、$F2$、$F3$、$F4$、$F5$ 作为自变量，生存状态指标 X_{Svl} 作为因变量，进行 Logistic 回归分析，具体步骤如下。

(1) Logistic 回归模型检验

本文对构建的 Logistic 回归模型进行了 Omnibus 检验、似然比检验(表 5)、Hosmer-Lemshow 检验(表 6)。在二元 Logistic 回归 Omnibus 检验中，$\chi^2=35.731$，$P<0.05$ 表示本次拟合的模型纳入的变量中，至少有一个变量的 OR 值有统计学意义。

由表 5 可知，对于模型的表现及有效性，显著性 P 值为 0.000，因而模型表现较好且有效。结果中 AIC 值为126.996，BIC 值为 146.253，相对较小，说明模型拟合效果较好。

表 5　似然比检验结果

Table 5　Results of likelihood ratio test

似然比卡方值	P	AIC	BIC
114.996	0.000***	126.996	146.253

表 6　Hosmer-Lemshow 检验结果

Table 6　Results of Hosmer-Lemshow test

步骤	卡方	自由度	显著性
1	6.178	8	0.627

注：*** 表示在 $P<0.001$ 水平下差异显著。

Hosmer-Lemshow 检验的显著性 P 值为 0.627＞0.05,认为数据中的信息已经被充分提取,故该模型拟合优度较高。

(2) 构建逻辑回归模型

本文构建的逻辑回归模型的参数结果如表 7 所示。

表 7 二元 Logistic 回归系数表

Table 7 Binary Logistic regression coefficients

实验组	回归系数	标准误差	Wald	P	OR	OR 值 95%置信区间	
						上限	下限
常数	−2.389	0.320	55.693	0.000***	0.092	0.049	0.172
癌症临床阶段因子	0.202	0.218	0.860	0.354	1.224	0.799	1.875
更年期状态因子	0.081	0.239	0.114	0.736	1.084	0.678	1.732
妊娠因子	−0.175	0.255	0.473	0.492	0.839	0.509	1.383
肿瘤转移扩散因子	1.394	0.277	25.334	0.000***	4.029	2.342	6.933
生长部位级别因子	0.232	0.240	0.932	0.334	1.261	0.787	2.019

注:*** 表示在 P＜0.001 水平下差异显著。

由表 7 可知,肿瘤转移扩散因子的显著性 P 值为 0.000,达到了极其显著的水平,拒绝原假设,表明肿瘤转移扩散因子与生存状态之间存在显著关联。

肿瘤转移扩散因子的主要影响因素包括最后一次随访时患者的肿瘤状况(Tmr)、癌细胞是否已经扩散到远离原发灶的其他部位(M)、淋巴结受累情况(N)。这些因素共同作用使得肿瘤转移扩散因子对生存状态产生显著影响。具体而言,肿瘤转移扩散因子每增加一个单位,生存状态为 0(即死亡)的概率比生存状态为 1(即存活)的概率高出 302.919%。这一发现意味着癌症的转移和扩散情况对于患者的生存状态具有重大影响[21]。一旦癌症发生转移和扩散,病人的死亡率会显著增加。因此,肿瘤转移扩散对于宫颈癌患者的生存状态具有重要影响,是宫颈癌患者治疗和预后评估的关键因子。

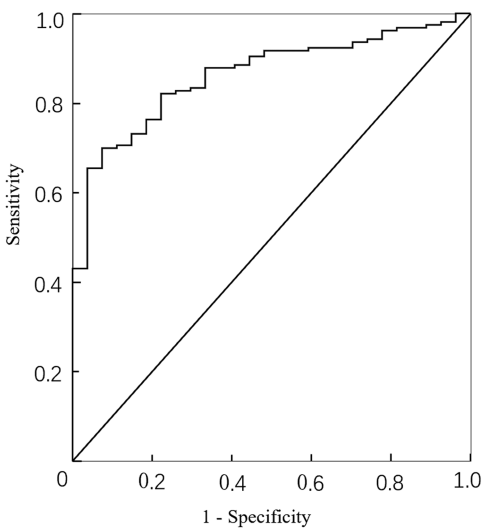

图 1 ROC 曲线图

Fig. 1 ROC curve

由 ROC 曲线图(图 1)可知,ROC 曲线下的面积值为 0.861,说明该模型在判别能力方面表现出良好的性能。为了进一步评估模型的性能,计算了该二元 Logistic 回归拟合指标,结果如表 8 所示。

表 8 二元 Logistic 回归拟合指标

Table 8 Binary Logistic regression fitting index

预测总精度	召回率	精确度	F_1	AUC
0.842	0.842	0.810	0.820	0.861

根据表 8 数据，模型预测总精度与召回率均达到了 84.2%，精确度为 81.0%，AUC 值为 0.861，综合考虑预测总精度、召回率、精确度和 F_1 分数，发现该模型在分类任务上表现良好，具有较高的预测能力和可靠性。

3　结论

本文基于宫颈癌临床 14 个指标进行统计分析，研究了宫颈癌这些临床特征的相关性、宫颈癌临床特征的基本结构，并且构建了基于宫颈癌临床特征的生存状态预测模型，得到如下结果：

(1) 基于 Spearman 相关性分析发现，患者所处癌症的具体临床阶段(Stg)与癌细胞是否已经扩散到远离原发灶的其他部位(M)、淋巴结受累情况(N)、原发肿瘤的分类(T)呈现相关关系，初次病理诊断年龄(Age)与女性更年期状态(Mpe)呈现显著相关关系，最后一次随访患者的肿瘤状态(Tmr)与癌细胞是否已经扩散到远离原发灶的其他部位(M)呈现相关关系，淋巴结受累情况(N)与原发肿瘤的分类(T)呈现相关关系，生育活产儿的成功妊娠量(Fty)与宫颈癌的临床阶段(Stg)正相关，低体重与宫颈癌较高的临床阶段相关。

(2) 通过因子分析寻求初次病理诊断年龄(X_{Age})、女性更年期状态(X_{Mpe})、怀孕总数(X_{Pgt})、生育活产儿的成功妊娠量(X_{Fty})、癌细胞是否已经扩散到远离原发灶的其他部位(X_{M})、淋巴结受累情况(X_{N})、原发肿瘤的分类(X_{T})、生长部位级别(X_{Gth})、最后一次随访患者的肿瘤状态(X_{Tmr})、病人所处癌症的具体临床阶段(X_{Stg})与吸烟程度(X_{Smk})11 个指标的基本结构，提取了癌症临床阶段因子、更年期状态因子、妊娠因子、肿瘤转移扩散因子和生长部位级别因子，可以从这 5 个方面来描述宫颈癌患者的临床特征，进而基于这 5 个因子构建生存状态预测模型。该方法简化了数据结构，提取的潜因子为后续研究提供了支持。

(3) 基于因子分析的二元 Logistic 回归结果显示，该模型具有较高的预测能力和可靠性。肿瘤转移扩散因子的显著性 P 值为 0.000，达到了极其显著的水平。癌症的转移和扩散情况对于患者的生存状态具有重大影响。一旦癌症发生转移和扩散，病人的死亡率会显著增加。因此，肿瘤转移扩散因子是宫颈癌患者治疗和预后评估的关键因子。肿瘤转移扩散因子的主要影响因素包括最后一次随访时患者的肿瘤状况(Tmr)、癌细胞是否已经扩散到远离原发灶的其他部位(M)、淋巴结受累情况(N)。研究结果具有重要的临床应用价值，这可以为临床医生提供更加准确的病情评估因子，有助于临床医生更精准有效地制定个体化治疗方案。

参考文献：

[1] BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2024, 74(3): 229-263.

[2] CHEN W Q, ZHENG R S, BAADE P D, et al. Cancer statistics in China, 2015[J]. CA: A Cancer Journal for Clinicians, 2016, 66(2): 115-132.

[3] YI X, LI J J, YU S, et al. A new PCR-based mass spectrometry system for high-risk HPV, part Ⅰ: Methods[J]. American Journal of Clinical Pathology, 2011, 136(6): 913-919.

[4] 周潇妮，唐旭秀，蔡丽萍，等. 新辅助化疗后机器人辅助与开腹手术治疗局部晚期子宫颈癌术后生存影响因素的对比分析[J]. 机器人外科学杂志(中英文)，2024，5(2)：178-185.

[5] DU H, YI J, WU R F, et al. A new PCR-based mass spectrometry system for high-risk HPV, part Ⅱ: Clinical trial[J]. American Journal of Clinical Pathology, 2011, 136(6): 920-923.

[6] 吴斯旻. 13 部门发文促进 HPV 疫苗接种业界建议优化接种政策[N]. 第一财经日报，2023-11-17(A02).

[7] 王慧珂，朱奕潼，丁晗玥，等. 在低卫生资源环境下消除宫颈癌：从全球视角到中国视角[J]. 中国妇幼卫生杂志，2023，14(5)：1-7.

[8] 辛雪焕. 不同 HPV 感染状态宫颈癌患者临床病理特征及预后分析[D]. 济南：山东大学，2023.

[9] SALAZAR Y,ZHENG X,BRUNN D,et al. Microenvironmental Th9 and Th17 lymphocytes induce metastatic spreading in lung cancer[J]. The Journal of Clinical Investigation,2020,130(7):3560-3575.
[10] ASADZADEH Z,MOHAMMADI H,SAFARZADEH E,et al. The paradox of Th17 cell functions in tumor immunity[J]. Cellular Immunology,2017,322:15-25.
[11] 张美琴,陈鸣之.年轻妇女子宫颈癌 174 例临床及预后分析[J].中华妇产科杂志,2003,38(11):689-693.
[12] RAJEEVAN M S,SWAN D C,NISENBAUM R,et al. Epidemiologic and viral factors associated with cervical neoplasia in HPV-16-positive women[J]. International Journal of Cancer,2005,115(1):114-120.
[13] 彭俊,黄勇.90 例年轻宫颈癌患者临床特征及术后复发和预后相关因素分析[J].实用妇产科杂志,2016,32(1):42-45.
[14] WU P,XIONG H G,YANG M,et al. Co-infections of HPV16/18 with other high-risk HPV types and the risk of cervical carcinogenesis:A large population-based study[J]. Gynecologic Oncology,2019,155(3):436-443.
[15] WRIGHT J D,MATSUO K,HUANG Y M,et al. Prognostic performance of the 2018 international federation of gynecology and obstetrics cervical cancer staging guidelines[J]. Obstetrics and Gynecology,2019,134(1):49-57.
[16] JEONG S Y,PARK H,KIM M S,et al. Pretreatment lymph node metastasis as a prognostic significance in cervical cancer:Comparison between disease status[J]. Cancer Research and Treatment:Official Journal of Korean Cancer Association,2020,52(2):516-523.
[17] KAHN R,BADINER N,NICHOLSON N,et al. Without mandatory high-risk HPV co-testing,are we inadequately screening for adenocarcinoma of the cervix? [J]. Gynecologic Oncology,2021,162(S1):S336-S337.
[18] MACIOS A,DIDKOWSKA J,WOJCIECHOWSKA U,et al. Risk factors of cervical cancer after a negative cytological diagnosis in Polish cervical cancer screening programme[J]. Cancer Medicine,2021,10(10):3449-3460.
[19] 孙君华,叶永生,徐小晶.中青年宫颈癌患者预后影响因素分析[J].医药论坛杂志,2024,45(6):637-640.
[20] SUN C Z,FENG Y E. EPDRNA:A model for identifying DNA-RNA binding sites in disease-related proteins [J]. The Protein Journal,2024,43(3):513-521.
[21] 于海青,顾佳欣,樊国梁.宫颈癌分子调控机制的关键基因分析[J].内蒙古大学学报(自然科学版),2021,52(1):49-58.

Prediction Analysis of Survival Status Based on Clinical Characteristics of 307 Cervical Cancer Patients

SUN Pengzhe,XÜ Chunjie,YAN Zuwei,FENG Yonge
(*College of Science*,*Inner Mongolia Agricultural University*,*Hohhot* 010018,*China*)

Abstract:By exploring the relationship between clinical characteristics and survival status of cervical cancer patients,and then constructing an accurate and reliable survival status prediction model,new ideas and methods can be provided for the prevention,diagnosis,and treatment of cervical cancer. Based on the clinical follow-up data of 307 cervical cancer patients from the TCGA database,Spearman correlation analysis is used to explore the correlation between the clinical features of cervical cancer,and innovatively employed factor analysis to investigate the basic structure of these clinical features. A binary logistic regression survival state prediction model based on factor analysis was proposed. The model's overall prediction accuracy was 84.2%,the recall rate was 84.2%,the precision was 81.0%,and the AUC value of the model was 0.861,indicating that the model performed well in classification tasks with high predictive capability and reliability.

Key words:cervical cancer; survival status; factor analysis; Logistic regression