论著·智能医学与预测模型

整合多模态信息的宫颈癌前病变风险预测模型:基于修正泊松回归的列线图开发与验证

杨 欢[1,3],晏玉玲[1,3],王延琨[2,3],李思婧[1,3],杨 君[1,3],龙 馨[1,3] [1]重庆医科大学附属妇女儿童医院,重庆市妇幼保健院妇产科,重庆;[2]重庆市妇幼疾病预防控制与公共卫生研究中心,重庆;[3]国家卫生健康委出生缺陷与生殖健康重点实验室,重庆

[摘要] **目的** 宫颈癌作为全球女性第四大常见恶性肿瘤,严重威胁全球女性健康,早期识别并干预高级别鳞状上皮内病变(high-grade squamous intraepithelial lesion,HSIL)对于宫颈癌及癌前病变的防控具有重要意义。目前的筛查手段较为单一,漏诊风险与过度诊疗并存,如何实现高级别鳞状上皮内病变(≥HSIL)的个体化精准诊疗效成为临床防控宫颈癌及癌前病变的重要问题,本研究旨在整合人口统计学、社会行为学、分子生物学及影像形态学等多模态信息,通过识别高级别病变的独立危险因素,构建基于修正泊松回归的个体化风险预测列线图模型,实现宫颈癌及癌前病变的精准诊疗和个体分层。**方法** 采用回顾性观察性研究设计,选取2022年7月至2024年7月在重庆医科大学附属妇女儿童医院宫颈疾病诊治中心接受活检的741例患者作为研究对象。依据组织病理学金标准将患者分为2组:低级别鳞状上皮内病变组(≤LSIL组)484例,包括慢性宫颈炎206例、LSIL 278例;高级别鳞状上皮内病变组(≥HSIL组)257例,包括HSIL 243例、宫颈原位腺癌8例、宫颈癌6例。系统收集包括年龄、孕产史、人乳头瘤病毒(human papilloma virus, HPV)疫苗接种史、液基细胞学(thinprep cytology test,TCT)结果、HPV分型、HPV持续感染时间及阴道镜印象(如醋白上皮厚度、血管形态等)在内的多模态信息。针对连续变量采用独立样本 t 检验(数据符合正态分布且方差齐性)、分类变量采用Pearson χ^2检验(理论频数<5时采用Fisher精确检验)进行单因素分析,在将单因素分析中 $P<0.1$ 的变量纳入候选集、并强制控制年龄、孕次及产次等混杂因素后,采用修正泊松回归(稳健方差)并使用后退法筛选变量进入多因素分析,随后应用多因素修正泊松回归确定独立危险因素并计算 RR 及其95%CI。利用R软件构建列线图模型,并通过AUC和校准曲线(Bootstrap法,B=500)评估模型的区分度与一致性。**结果** 单因素分析显示,年龄($t=-2.264$,$P=0.024$)、孕次($t=-3.527$,$P<0.001$)、产次($t=-3.222$,$P<0.001$)、避孕方式($\chi^2=10.712$,$P=0.013$)、TCT结果($\chi^2=44.883$,$P<0.001$)、HPV分型($\chi^2=15.692$,$P=0.001$)、HPV累积感染时间($\chi^2=11.670$,$P=0.009$)、HPV疫苗接种情况($\chi^2=15.685$,$P<0.001$)、阴道镜印象($\chi^2=87.284$,$P<0.001$)、醋白情况($\chi^2=72.771$,$P<0.001$)、血管形态(点状血管/镶嵌血管)($\chi^2=67.099$,$P<0.001$)以及宫颈转化区类型($\chi^2=20.316$,$P<0.001$)在2组间差异均具有统计学意义($P<0.05$)。多因素修正泊松回归分析识别出5项独立预测指标:①年龄($RR=1.015$,95%CI:1.000~1.030,$P=0.048$);②孕次($RR=1.066$,95%CI:1.002~1.134,$P=0.043$);③HPV持续感染时间≥1年($RR=1.405$,95%CI:1.010~1.956,$P=0.044$);④阴道镜印象为≥HSIL($RR=2.108$,95%CI:1.545~2.878,$P<0.001$);⑤HPV疫苗接种(保护因素,$RR=0.631$,95%CI:0.510~0.781,$P<0.001$)。基于上述因素构建的列线图模型预测高级别病变风险的AUC为0.740(95%CI:0.702~0.778)。校准曲线显示预测概率与实际观察值具有良好一致性,校准斜率为0.958 0,Brier得分为0.191 4。**结论** 本研究整

[基金项目] 肿瘤侵袭转移教育部重点实验室开放基金(2024KFKT012);2022年国家临床重点专科建设项目(妇产科)

[通信作者] 龙 馨,E-mail:4261749@qq.com

2097-0927/Copyright © 2026, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循 CC BY 许可协议 (https://creativecommons.org/licenses/by/4.0/)

引用格式:杨欢,晏玉玲,王延琨,等. 整合多模态信息的宫颈癌前病变风险预测模型:基于修正泊松回归的列线图开发与验证[J]. 陆军军医大学学报,2026,48(07):905-913. DOI:10.16016/j.2097-0927.202512050.

Citation: YANG H, YAN Y L, WANG Y K, *et al*. Risk prediction model for precancerous cervical lesions by integrating multimodal data: development and validation of a nomogram based on modified Poisson regression[J]. J Army Med Univ, 2026, 48(07): 905-913. DOI: 10.16016/j.2097-0927.202512050.

合了多模态信息构建列线图模型,具有良好的区分度与临床信度。可以直观化、可视化实现个体化风险量化,有助于优化宫颈癌防控效率。

［关键词］ 宫颈鳞状上皮内病变;信息整合;风险因素;修正泊松回归;列线图
［中图法分类号］ R195.4;R730.1;R737.33 ［文献标志码］ A

Risk prediction model for precancerous cervical lesions by integrating multimodal data: development and validation of a nomogram based on modified Poisson regression

YANG Huan[1,3], YAN Yuling[1,3], WANG Yankun[2,3], LI Sijing[1,3], YANG Jun[1,3], LONG Xin[1,3] [1]Department of Obstetrics and Gynecology, Women and Children's Hospital of Chongqing Medical University (Chongqing Health Center for Women and Children), Chongqing; [2]Chongqing Research Center for Maternal and Child Disease Prevention, Control and Public Health, Chongqing; [3]Key Laboratory of National Health Commission for Birth Defects and Reproductive Health, Chongqing, China

［**Abstract**］ **Objective** Cervical cancer, as the fourth most common malignancy among women worldwide, poses a serious threat to global women's health. Early identification and intervention of high-grade squamous intraepithelial lesions (HSIL) are of great significance for the prevention and control of cervical cancer and precancerous lesions. Current screening methods are relatively singular, leading to a coexistence of missed diagnosis risk and overtreatment. How to achieve individualized precision diagnosis and treatment for lesions staged ≥HSIL has become a vital clinical issue. This study aims to integrate multimodal information, including demographics, social behavior, molecular biology, and radiology, identify independent risk factors for high-grade lesions, and construct an individualized risk prediction nomogram model based on modified Poisson regression (MPR) to achieve precision diagnosis and treatment and individual stratification of cervical cancer and precancerous lesions. **Methods** A retrospective observational study was conducted on 741 patients who underwent biopsies at Cervical Disease Diagnosis and Treatment Center of Women and Children's Hospital of Chongqing Medical University between July 2022 and July 2024. Based on the histopathological gold standards, they were divided into a low-grade squamous intraepithelial lesion group (≤LSIL group, n=484, including 206 cases of chronic cervicitis and 278 cases of LSIL) and a high-grade squamous intraepithelial lesion group (≥HSIL group, n=257, including 243 cases of HSIL, 8 cases of cervical adenocarcinoma *in situ*, and 6 cases of cervical cancer). Multimodal data were systematically collected, including age, parity, human papillomavirus (HPV) vaccination history, thin-prep cytologic test (TCT) results, HPV genotypes, duration of persistent HPV infection, and colposcopic impression (such as acetowhite epithelial thickness and vascular patterns). Independent samples t-test (for continuous variables with normal distribution and homogeneity of variance) and Pearson *Chi*-square test (Fisher's exact test when theoretical frequency <5) were used for univariate analysis. After incorporating variables with P<0.1 from univariate analysis into the candidate set and after forcibly controlling confounders such as age, gravidity and parity, a backward elimination method was used within the MPR framework to identify independent risk factors and calculate risk ratios (*RR*) with 95%*CI*. A nomogram was constructed using *R* software, and its discrimination and consistency were evaluated *via* the AUC value of ROC curve analysis and calibration curve analysis (Bootstrap method, B=500). **Results** Univariate analysis showed statistically significant differences between the 2 groups in terms of (t=−2.264, P=0.024), gravidity (t=−3.527, P<0.001), parity (t=−3.222, P<0.001), contraceptive methods (χ^2=10.712, P=0.013), TCT results (χ^2=44.883, P<0.001), HPV genotypes (χ^2=15.692, P=0.001), cumulative duration of HPV infection (χ^2=11.670, P=0.009), HPV vaccination status (χ^2=15.685, P<0.001), colposcopic impression (χ^2=87.284, P<0.001), acetowhite findings (χ^2=72.771, P<0.001), vascular patterns (punctation/mosaicism) (χ^2=67.099, P<0.001), and type of cervical transformation zone (χ^2=20.316, P<0.001). Multivariate MPR identified 5 independent predictors: ① age (*RR*=1.015,

95%*CI*: 1.000 to 1.030, *P*=0.048); ② gravidity (*RR*=1.066, 95%*CI*: 1.002 to 1.134, *P*=0.043); ③ duration of persistent HPV infection (*RR*=1.405, 95%*CI*: 1.010 to 1.956, *P*=0.044); ④ colposcopic impression (*RR*=2.108, 95%*CI*: 1.545 to 2.878, *P*<0.001); and ⑤ HPV vaccination (protective factor, *RR*=0.631, 95%*CI*: 0.510 to 0.781, *P*<0.001). The nomogram model constructed based on these factors achieved an AUC value of 0.74 (95%*CI*: 0.702 to 0.778) in predicting ≥HSIL risk. The calibration curve showed good consistency between predicted probabilities and actual observed values, with a calibration slope of 0.958 0 and a Brier score of 0.191 4. **Conclusion** Our nomogram model constructed by integrating multimodal information demonstrates good discrimination and clinical reliability. It enables intuitive and visualized quantification of individualized risk, facilitating optimized efficiency in cervical cancer prevention and control.

[**Key words**] squamous intraepithelial lesions of the cervix; information integration; risk factors; modified Poisson regression; nomograms

Supported by the Open Fund of Key Laboratory of Ministry of Education for Tumor Invasion and Metastasis (2024KFKTO12), and the National Clinical Key Specialty Construction Program in 2022 (Obstetrics and Gynecology). Corresponding author: LONG Xin, E-mail: 4261749@qq.com

宫颈癌是全球女性常见的恶性肿瘤之一,全球癌症观察站(Global Cancer Observatory, GLOBOCAN)2022年的数据显示,全球约有66万新发宫颈癌病例和35万死亡病例,是女性癌症相关死亡的主要原因之一[1-2]。持续性高危人乳头瘤病毒感染(human papilloma virus,HPV)是宫颈高级别病变发生发展的主要因素,通过高效的筛查可以识别高级别鳞状上皮内病变(high-grade squamous intraepithelial lesion,HSIL),对HSIL进行干预是宫颈疾病防控的核心窗口[3]。尽管早期筛查与推广疫苗接种在总体上降低了疾病负担,但当前防控体系仍面临两大挑战:①如何在保证安全性的前提下更早、更精准地识别HSIL高风险患者;②如何在分级诊疗中优化宫颈活检指征,平衡过度检查与漏诊风险。基于风险的管理理念正逐步取代单一阈值式策略,而其有效实施依赖于多源信息的充分整合与预测模型的建立[4]。

既往研究认为,年龄、首次性行为年龄、HPV感染、病毒载量及细胞学异常等是HSIL的独立风险因素[5-6]。但宫颈病变的发生发展还与人口统计学、社会行为学特征、分子生物学检测、影像形态学特征(阴道镜印象)、HPV持续感染状态及时间累积风险等多重风险因素的影响有关[7]。传统的预测模型仅基于单一模态(如HPV基因分型、细胞学或阴道镜印象)进行风险评估,存在主观性强、外部可推广性差、对HPV感染累积时间等动态信息利用不足等局限[4,8]。因此,亟须建议融合上述多模态信息的综合预测模型,以提高风险识别的准确性与临床可用性[8]。

本研究整合人口统计学(年龄、孕产史、绝经状态等)、社会行为学(吸烟史、避孕方式、HPV疫苗接种情况等)、分子生物学检测[液基细胞学(thinprep cytology test,TCT)结果、HPV分型、HPV单一/多重感染等]、影像形态学(阴道镜印象,如醋白上皮厚度、血管形态、转化区类型、病变面积等)及HPV累积感染时间等多模态信息,进行回顾性分析,识别HSIL风险因素,为宫颈癌前病变的精准风险评估与个体化干预提供科学依据。

1 资料与方法

1.1 一般资料

采用回顾性观察性研究设计,选取2022年7月至2024年7月在重庆医科大学附属妇女儿童医院宫颈疾病诊治中心进行活检患者741例,依据组织病理学金标准将患者分为2组:低级别鳞状上皮内病变组(≤LSIL组)484例;高级别鳞状上皮内病变组(≥HSIL组)257例。纳入标准:①19~75岁;②具有明确的宫颈癌筛查结果及宫颈组织学诊断结果;③具有保存完整且评估充分的阴道镜图像,由2名高年资妇科医师审核;④既往无宫颈病变及宫颈恶性肿瘤病史;⑤无其他恶性肿瘤病史;⑥非哺乳期、非妊娠期女性。排除标准:①一般资料、宫颈癌筛查、病理资料不完整;②阴道镜图像采集不完整、评估不充分;③有子宫颈手术史。本研究经重庆医科大学附属妇女儿童医院(重庆市妇幼保健院)医学伦理委员会批准[(2025)伦审(科)059号],采取封闭式问卷建档信息资料:包括基本信息、月经史、婚育史、避孕史、既往史、家族史、性行为、吸烟史等。共纳入符合标准的患者741例。

1.2 资料收集方法

回顾性分析患者电子病历、阴道镜印象及电话随访资料,系统整合多模态信息。共构建5个数据模态:①人口统计学模态(年龄、孕产史、绝经状态等);②社会行为学模态(吸烟史、避孕方式、HPV疫苗接种情况等);③分子生物学模态(TCT结果、HPV分型、HPV单一/多重感染等);④影像形态学模态(阴道镜下醋白上

皮厚度、血管形态、转化区类型、病变面积及阴道镜印象等);⑤动态演变模态(HPV累积感染时间)。

1.2.1 宫颈细胞学检查 由重庆医科大学附属妇女儿童医院病理科医师采用TCT制片、阅片及诊断,诊断标准按照2014版贝塞斯达系统(The Bethesda System, TBS)细胞学诊断标准。

1.2.2 HPV检测 HPV分型检测采用荧光PCR法,使用人乳头瘤病毒核酸分型检测试剂盒,可检测23种人乳头瘤病毒亚型,包括13种高危型表达(HPV-16、18、31、33、35、39、45、51、52、56、58、59、68)、4种中危型(HPV-53、66、73、82)和6种低危型(HPV-6、11、42、43、44、81)。其中HPV-16和HPV-18是包括子宫颈、肛门、阴道、阴茎、外阴和口咽在内的多种癌症进展的主要原因[9-10]。本研究结果显示,HPV-16、18、52、58是重庆地区宫颈病变患者中最常见的高危亚型,这与重庆地区既往HPV流行病学特征一致[11]。

1.2.3 阴道镜检查 由我院阴道镜医师执行,遵照2019年ASCCP阴道镜诊断标准[12]。

1.2.4 组织病理学检查 以WHO(2020)女性生殖肿瘤分类第5版为诊断标准。若同一患者子宫颈不同部位病检结果性质不同时,按病变性质最严重的结果进行统计。

1.3 统计学分析

采用SPSS 27.0软件进行统计学分析。先评估变量缺失情况,缺失率<5%者采用链式方程多重插补法填补,缺失率≥5%的变量予以剔除。年龄、孕次、产次采用独立样本t检验,TCT结果、HPV分型、阴道镜印象等分类变量采用Pearson χ^2检验(理论频数<5时用Fisher精确检验),以P<0.1为标准筛选单因素分析中有统计学意义的变量进入多因素分析。采用多因素修正泊松回归识别≥HSIL的独立风险因素。采用修正泊松回归稳健方差估计*RR*及其95%*CI*。模型采用后退法筛选变量(剔除标准P>0.05),并控制年龄、孕产次等混杂因素,输出*RR*及95%*CI*。最终多因素回归模型纳入5个独立变量,利用R软件rms包构建列线图模型,以定量预测≥HSIL风险并赋予各预测因子权重,为临床个性化风险评估与干预提供工具。P<0.05为差异具有统计学意义。

2 结果

2.1 一般资料及病理结果

共纳入患者741例,年龄19~75岁,平均年龄为(39.04±11.47)岁,病理诊断结果显示:≤LSIL组484例(包括慢性宫颈炎206例、LSIL 278例),≥HSIL组257例(包括HSIL 243例、宫颈原位腺癌8例、宫颈癌6例)。

2.2 纳入研究对象的临床特征及单因素分析

描述性分析结果显示,年龄、孕次、产次、避孕方式、TCT结果、HPV分型、HPV累积感染时间、HPV疫苗接种情况、阴道镜印象、醋白情况、血管形态(点状血管/镶嵌血管)以及宫颈转化区类型在2组间差异具有统计学意义(P<0.05,表1)。

表1 纳入研究患者的临床特征及组间比[例(%),$\bar{x}\pm s$]

Table 1 Clinical characteristics of the participants and intergroup comparisons [n(%),$\bar{x}\pm s$]

Risk factors	≤LSIL (n=484)	≥ HSIL(n=257)	Test statistic	P
Age/years	38.34±11.25	40.34±11.80	−2.264	0.024
Gravidity/times	2.26±1.83	2.77±1.97	−3.527	<0.001
Parity /times	0.97±0.85	1.20±0.96	−3.222	<0.001
Clinical symptoms			0.027	0.870
Yes	47(9.71)	24(9.34)		
No	437(90.29)	233(90.66)		
Smoking status			1.416	0.234
Yes	67(13.84)	44(17.12)		
No	417(86.16)	213(82.88)		
Contraceptive methods			10.712	0.013
Condom	286(59.09)	128(49.81)		
Oral contraceptives	12(2.48)	2(0.78)		
Intrauterine device	58(11.98)	34(13.23)		
Others (coitus interruptus, sterilization, none)	128(26.45)	93(36.19)		
Menopausal status			0.443	0.506
Yes	90(18.60)	53(20.62)		
No	394(81.40)	204(79.38)		

续表1 纳入研究患者的临床特征及组间比[例(%),$\bar{x}\pm s$]

Table 1(continued) Clinical characteristics of the participants and intergroup comparisons [n(%),$\bar{x}\pm s$]

Risk factors	≤LSIL (n=484)	≥ HSIL(n=257)	Test statistic	P
Duration of menopause			4.439	0.218
Premenopausal	394(81.40)	204(79.38)		
≤5 years	45(9.30)	20(7.78)		
>5 to <15 years	36(7.44)	22(8.56)		
≥15 years	9(1.86)	11(4.28)		
TCT results			44.883	<0.001
NILM	255(52.69)	104(40.47)		
ASC-US	99(20.45)	51(19.84)		
LSIL	19(3.93)	14(5.45)		
ASC-H	90(18.60)	44(17.12)		
HSIL	15(3.10)	42(16.34)		
AGC	6(1.24)	2(0.78)		
HPV genotyping			15.692	0.001
High-risk single infection (16, 18, 52, 58)	198(40.91)	105(40.86)		
Other single type infection	131(27.07)	41(15.95)		
Multiple infections	142(29.34)	105(40.86)		
Negative	13(2.69)	6(2.33)		
Duration of infection			11.670	0.009
Non-persistent infection	13(2.69)	6(2.33)		
<1 years	259(53.51)	164(65.34)		
1 to<2 years	175(36.16)	62(24.70)		
≥2 years	37(7.64)	25(9.96)		
HPV vaccination			15.685	<0.001
Yes	84(17.36)	77(29.96)		
No	400(82.64)	180(70.04)		
Colposcopic impression			87.284	<0.001
Normal colposcopy	72(14.88)	29(11.28)		
LSIL	340(70.25)	111(43.19)		
HSIL	71(14.67)	108(42.02)		
Suspicious for carcinoma	1(0.20)	9(3.51)		
Acetowhitening			72.771	<0.001
Absent	57(11.78)	26(10.12)		
Thin	361(74.59)	125(48.64)		
Thick	66(13.63)	106(41.24)		
Punctation/mosaicism			67.099	<0.001
Absent	399(82.44)	152(59.14)		
Fine	70(14.46)	56(21.79)		
Coarse	15(3.10)	49(19.07)		
Lesion area/cervical area			4.122	0.127
<1/3	90(18.60)	58(22.57)		
1/3 to 2/3	21(4.34)	5(1.95)		
>2/3	373(77.07)	194(75.49)		
Transformation zone types			20.316	<0.001
Type Ⅰ	256(52.89)	96(38.35)		
Type Ⅱ	186(38.43)	118(45.91)		
Type Ⅲ	42(8.68)	43(16.73)		

NILM: negative for intraepithelial lesion or malignancy; ASC-US: atypical squamous cells of undetermined significance; ASC-H: atypical squamous cells, cannot exclude HSIL; LSIL: low-grade squamous intraepithelial lesion; HSIL: high-grade squamous intraepithelial lesion; AGC: atypical glandular cells; TCT: thin-prep cytologic test

2.3 多因素修正泊松回归分析

在将单因素分析中$P<0.1$的变量纳入候选集、并强制控制年龄、孕次及产次等混杂因素后,采用修正泊松回归(稳健方差)并使用后退法筛选变量。最终显示:年龄、孕次、HPV持续感染时间及阴道镜印象为≥HSIL的独立危险因素($P<0.05$),HPV疫苗接种为保护因素($P<0.001$)。具体而言:①年龄每增加1岁,≥HSIL发生风险增加($\beta=0.015$, *SE*. robust =0.007,*RR*=1.015,95%*CI*:1.000~1.030);②孕次每增加1次,≥HSIL发生风险增加($\beta=0.064$, *SE*. robust =0.031,*RR*=1.066,95%*CI*:1.002~1.134);③HPV持续感染时间≥1年者较感染<1年者风险升高($\beta=0.340$, *SE*. robust =0.169,*RR*=1.405,95%*CI*:1.010~1.956);④阴道镜印象为≥HSIL者发生≥HSIL的风险是阴道镜印象为≤LSIL者的($\beta=0.746$, *SE*. robust=0.159, *RR*=2.108, 95%*CI*:1.545~2.878),为最强预测因子;⑤接种HPV疫苗可使≥HSIL发生风险降低($\beta=-0.460$, *SE*. robust=0.109,*RR*=0.631,95%*CI*:0.510~0.781)。另外,TCT结果、避孕方式、宫颈转化区类型等变量虽在单因素分析中提示相关性,但在多因素候选集中经后退法筛选后未进入最终模型,提示其效应可能被阴道镜印象、HPV持续感染等更强预测因子所解释,或存在变量间信息重叠(表2)。

2.4 列线图预测模型

2.4.1 模型的构建 列线图预测模型根据多因素修正泊松回归分析筛选风险因素和保护因素建立(见图1)。最终纳入5个独立变量,各变量赋予相应权重,将5项指标具体分值相加后得出总分,对应的概率线即可得出该患者发生HSIL的具体概率值。本研究样本量充足,结局事件(≥HSIL)257例,结局事件数与变量数比值为51.4,远高于10~15个事件/变量的多因素回归最低标准及20~30个事件/变量的理想标准,可确保模型拟合稳定、结果可靠。

2.4.2 模型的准确性评估 预测模型的ROC曲线见图2,AUC为0.740(95%*CI*:0.702~0.778),表明该模型具备中等偏上区分能力,可有效支撑临床高、低风险人群分流。该模型以约74%的概率将阳性样本排在阴性样本之前,体现出具有实际参考价值的预测效能。

表2 宫颈高级别病变的多因素回归分析结果(n=257)

Table 2 Multivariate regression analysis results for high-grade cervical lesions(n=257)

Risk factors	β	*SE*. robust	*RR*(95%*CI*)	*P*
Gravidity	0.064	0.031	1.066(1.002—1.134)	0.043
Age	0.015	0.007	1.015(1.000—1.030)	0.048
HPV vaccination	−0.460	0.109	0.631(0.510—0.781)	<0.001
Colposcopic impression	0.746	0.159	2.108(1.545—2.878)	<0.001
Duration of persistent HPV infection	0.340	0.169	1.405(1.010—1.956)	0.044

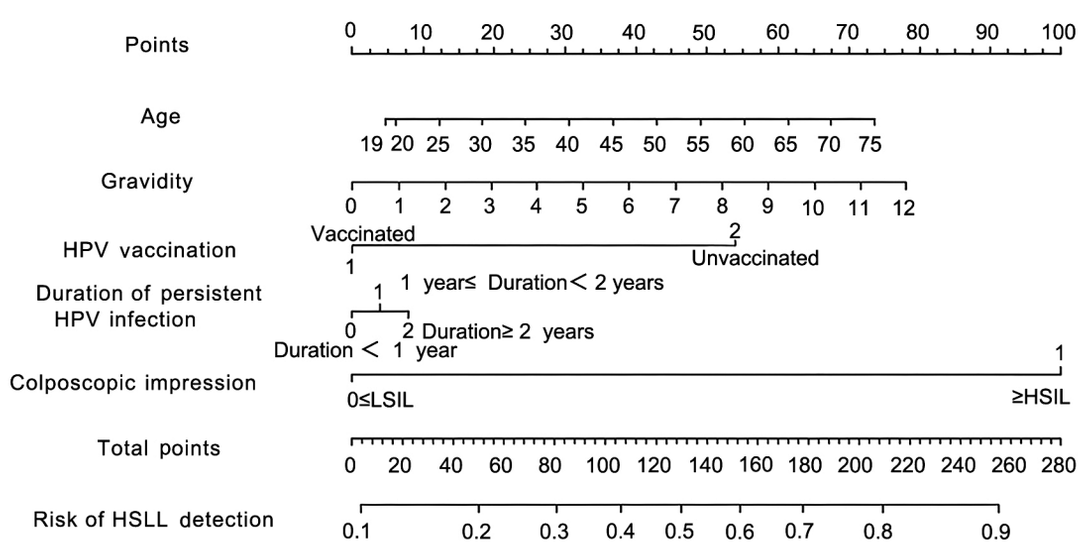


图1 预测宫颈高级别病变因素的列线图预测模型

Figure 1 The nomogram prediction model for high-grade cervical lesions

模型校准能力评估见图3。校准曲线显示表观曲线(Apparent)与Bootstrap偏倚校正曲线(Bias-corrected)整体接近理想参考线(Ideal),提示预测概率与实际观察风险具有较好一致性。Bootstrap内部验证(B=500)进一步量化显示:校准截距为-0.016 8(接近0),校准斜率为0.958 0(接近1),提示系统性偏差很小且过拟合程度较低;Brier得分为0.191 4,表明模型总体概率预测误差处于可接受水平。结合图3可见,中等风险区间曲线略高于理想线,提示该区间可能存在轻微低估风险趋势,但偏离幅度较小且经偏倚校正后仍保持良好一致性。

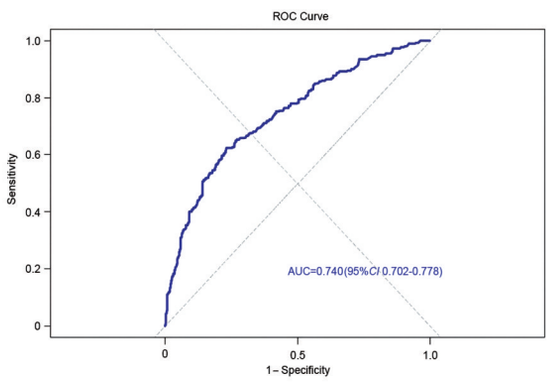


图2　预测宫颈高级别病变因素列线图预测模型的ROC曲线

Figure 2　Receiver operating characteristic (ROC) curve of the nomogram prediction model for high-grade cervical lesions

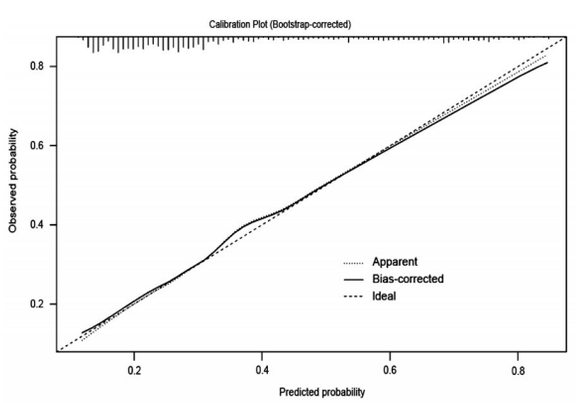


图3　预测宫颈高级别病变因素列线图预测模型的校准曲线

Figure 3　Calibration curve of our nomogram prediction model for high-grade cervical lesions

2.4.3　列线图模型的应用　　列线图模型可直观量化患者≥HSIL风险,为基层医疗机构"筛查-治疗"策略提供便捷工具。

高风险案例:患者42岁,孕5次,无生育要求,未接种HPV疫苗,HPV16型持续感染2.5年,阴道镜印象HSIL。模型评分为32.5+32.5+53+7.5+100=225.5分,预测风险约为84%。针对该患者,可优先选择宫颈锥切术,避免多阶段转诊延误,降低病变进展风险[4,13]。

低风险案例:患者30岁,孕1次,有生育要求,完成HPV疫苗接种,HPV感染6个月,阴道镜印象LSIL。模型评分17.5+5+0+0+0=22.5分,预测风险约为14%。针对该患者,可行常规随访,12个月后复查HPV+TCT,基层无需额外增加诊疗资源,避免过度医疗。

综上,列线图模型可以通过精准风险分层,对高风险患者进行有效干预,优化低风险患者的管理流程,提高宫颈高级别病变的防控效率。

3　讨论

3.1　整合多模态数据可有效破解单一筛查维度的局限性

宫颈癌前病变的精准风险分层是降低宫颈癌发病率和死亡率的核心举措,当前临床依赖单一模态的风险评估模式,因主观性强、动态信息利用不足等局限,难以平衡漏诊风险与过度医疗矛盾,亟须多源信息整合的系统化预测工具破解困局[14]。

与传统仅依赖TCT和HPV检测的预测模型相比,本研究的核心优势为三维创新:①维度创新,突破单一模态局限,将HPV持续感染时间这一动态风险因子与阴道镜影像形态学特征、疫苗接种状态等关键变量系统整合,捕捉病变进展的多维度驱动机制,使风险评估更贴合疾病自然史规律;②效能创新,预测模型AUC达0.740,具备中等偏上的风险区分能力,可精准量化个体发病概率,且所有纳入指标为基层医疗机构易获取项,无需额外增加检测成本,完美适配分级诊疗场景;③应用创新,模型可直接指导临床决策,为宫颈癌防控提供了可及性强、操作性高的循证工具。

3.2　阴道镜印象与HPV持续感染时间是≥HSIL发生的核心风险锚点

本研究通过量化分析发现,阴道镜印象是模型中贡献度最高的风险因素(*RR*=2.108),处于病变评估的核心地位;HPV持续感染时间每增加1年,发病风险上升约10%,其中持续感染≥1年者*RR*达1.405,为制定个体化随访间隔提供了直接数据支撑,例如将该人群随访周期缩短至6个月;而HPV疫苗接种表现出显著的保护效应(*RR*=0.631),进一步印证了"疫苗接种+精准筛查"联合防控策略的科学依据。

3.3 列线图模型在辅助分级诊疗中兼具可解释性与操作性

传统回归模型为方程式,如Cox回归模型、Logistic回归模型等,每个变量的系数都具有明确的临床意义,模型的可解释性强,但其预测值需要进行公式换算[15-16],不适用于繁琐的临床工作。深度学习模型(deep learning model, DL model),如卷积神经网络(convolutional neural network, CNN)及其变体,通过多层复杂的非线性变换来学习和提取图像特征,具有处理复杂、多维数据(如医学影像、基因组学数据)的能力,在某些预测任务中可能实现更高的AUC值和预测性能[14,17],但其决策过程的不透明性常导致诊断结果缺乏可解释性[18]。而列线图模型基于传统回归模型构建,不仅延续了传统回归模型可解释性强的优点,更能直观透明地辅助临床决策[19-21]。即使缺乏复杂统计模型或深度学习算法的专业知识,临床医师仍可根据列线图中风险因素对应的分数相加得出总分判断某患者发生≥HSIL的风险,快速给予临床建议,实现≥HSIL风险个体化预测及精准医疗,便于医患沟通,且基层医疗场景中易于推广应用。

3.4 研究局限性与未来多中心前瞻性验证的方向

本研究的局限性在于仅为单中心回顾性研究,可能存在选择偏倚。主要包括:①转诊中心偏倚,由于所采集病例均来源于本院宫颈疾病诊治中心就诊的患者,其病情复杂程度及病变风险可能高于普通人群。因此,本研究所得出的相关结论可能无法广泛适用于普通健康人群或患者群体。②阴道镜指征偏倚,纳入本研究的病例均为符合转诊阴道镜指征的患者,缺乏普通健康筛查人群,导致样本集中于宫颈病变高风险群体。③信息偏倚,本研究中HPV持续感染时间的确定主要依赖既往筛查记录,可能因为病历不完整、患者记忆偏差等产生测量误差,且模型构建未充分整合p16/Ki-67、甲基化等关键分子标志物,未能完全释放“分子-影像-临床”多模态数据整合的预测潜力,限制了模型对宫颈高级别病变风险进行更深层次、更精准评估的能力。

在后续研究可开展大规模、多中心、前瞻性的验证研究。在不同地域、不同级别医疗机构中收集数据,评估模型在异质性人群和多样化临床环境中的表现,同时纳入p16/Ki-67、甲基化等分子标志物[22],提升模型的普适性与稳定性,为宫颈高级别病变的风险预测提供更可靠的临床决策支持。

本研究构建的预测模型提供了一套标准化、可视化的风险评估工具,能精准实现宫颈癌的风险分层,缩小不同地区诊疗水平差距,均衡优质医疗资源,可以帮助临床优化宫颈癌干预策略,提升宫颈癌防控效率,降低疾病负担。

综上,本研究以多模态数据整合为核心突破点,以临床实用化为导向,构建的列线图预测模型虽仍有优化空间,但已实现了宫颈高级别病变风险评估的精准化、个体化与可及化。该研究不仅为临床精准分流与个体化干预提供了新型循证工具,更开创了宫颈癌前病变风险管理的新范式,为我国分级诊疗背景下宫颈癌防控策略的优化,以及全球宫颈癌消除目标的实现注入了强劲动力。

利益冲突 本文所有作者声明不存在利益冲突

参考文献:

[1] FILHO A M, LAVERSANNE M, FERLAY J, *et al*. The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide[J]. Int J Cancer, 2025, 156(7): 1336-1346. DOI:10.1002/ijc.35278.

[2] 张哲, 吴杰, 伍婷婷, 等. 基于RNA结合蛋白构建宫颈癌预后模型并实验验证[J]. 空军军医大学学报, 2024, 45(7): 805-812. DOI:10.13276/j.issn.2097-1656.2024.07.014.
ZHANG Z, WU J, WU T T, *et al*. Construction and validation of cervical cancer prognosis model based on RNA-binding proteins[J]. J Air Force Med Univ, 2024, 45(7): 805-812. DOI:10.13276/j.issn.2097-1656.2024.07.014..

[3] ZHANG Y L, PAN H, ZHAO X X, *et al*. The impact of HPV/HBV coinfection on cervical cancer risk: potential mediating role of Ki67[J]. J Med Virol, 2025, 97(7): e70454. DOI: 10.1002/jmv.70454.

[4] PERKINS R B, GUIDO R S, CASTLE P E, *et al*. 2019 ASCCP risk-based management consensus guidelines for abnormal cervical cancer screening tests and cancer precursors[J]. J Low Genit Tract Dis, 2020, 24(2): 102-131. DOI:10.1097/LGT.0000000000000525.

[5] YAMAGUCHI M, SEKINE M, HANLEY S J B, *et al*. Risk factors for HPV infection and high-grade cervical disease in sexually active Japanese women[J]. Sci Rep, 2021, 11(1): 2898. DOI:10.1038/s41598-021-82354-6.

[6] VARESANO S, CICCARESE G, PAUDICE M, *et al*. Evaluating HPV viral load and multiple infections for enhanced cervical cancer risk-based assessment[J]. Life (Basel), 2025, 15(2): 153. DOI:10.3390/life15020153.

[7] KUMAR R, DUBEY N, JAIN S. Leveraging machine learning for early detection of cervical cancer: analyzing demographic, clinical, and lifestyle risk factors[J]. Romanian J Biophys, 2025, 35(1): 33-40. DOI: 10.59277/rjb.2025.1.03.

[8] MADATHIL S, DHOUIB M, LELONG Q, *et al*. A multimodal deep learning model for cervical pre-cancers and cancers prediction: Development and internal validation study [J]. Comput Biol Med, 2025, 186: 109710. DOI:10.1016/j.compbiomed.2025.109710.

[9] XU H F, LIU Y, LUO Y L, *et al*. The risk stratification for cervical cancer and precursors of domestic HPV testing with HPV 16/18 genotyping in women with NILM cytology in CentralChina: a cohort study [J]. Front Oncol, 2021, 11: 716762. DOI:10.3389/fonc.2021.716762.

[10] DONG Z, HU R J, DU Y, *et al*. Immunodiagnosis and immunotherapeutics based on human papillomavirus for HPV-induced cancers [J]. Front Immunol, 2020, 11: 586796. DOI:10.3389/fimmu.2020.586796.

[11] DING H L, JIN H, TANG Y S. HPV genotype distribution and cervical lesions in Chongqing: a comprehensive analysis of 229, 770 females (2015-2023)[J]. BMC Infect Dis, 2025, 25(1): 760. DOI:10.1186/s12879-025-11128-6.

[12] PERKINS R B, GUIDO R S, CASTLE P E, *et al*. 2019 ASCCP risk-based management consensus guidelines for abnormal cervical cancer screening tests and cancer precursors[J]. J Low Genit Tract Dis, 2020, 24(2): 102-131. DOI:10.1097/lgt.0000000000000525.

[13] ABU-RUSTUM N R, YASHAR C M, AREND R, *et al*. NCCN guidelines® insights: cervical cancer, version 1.2024: featured updates to the NCCN guidelines[J]. J Natl Compr Cancer Netw, 2023, 21(12): 1224-1233. DOI: 10.6004/jnccn.2023.0062.

[14] JEMANE K C, KUYU M B, TEGENAW G S. Developing multimodal cervical cancer risk assessment and prediction model based on LMIC hospital patient card sheets and histopathological images[J]. BMC Med Inform Decis Mak, 2025, 25(1): 322. DOI:10.1186/s12911-025-03174-6.

[15] ELZEIN I M, CHAMSEDDINE A, ELTANBOLY A, *et al*. Cervical cancer perceived risk factors behavior using logistic regression technique[J]. J Biomed Res, 2025: 1-12. DOI: 10.7555/JBR.39.20250047.

[16] KOLASSERI A E, VENKATARAMANA B. Comparative study of machine learning and statistical survival models for enhancing cervical cancer prognosis and risk factor assessment using SEER data[J]. Sci Rep, 2024, 14(1): 22203. DOI:10.1038/s41598-024-72790-5.

[17] DAYALANE S, MURUGESAN S, MATHIVANAN S K, *et al*. Cervical cancer classification using deep learning approach using colposcopy images[J]. Neural Process Lett, 2025, 57(4): 65. DOI:10.1007/s11063-025-11770-w.

[18] SWARNKAR B, KHARE N, GYANCHANDANI M. Systematic review of deep learning techniques for gynecological cancer diagnosis [J]. IEEE Access, 2025, 13: 154754-154774. DOI:10.1109/ACCESS.2025.3603458.

[19] LIAO Y Y, HUANG Q D, SHEN G Q, *et al*. Molecular subtypes and nomogram for predictingthe prognosisof cervical cancer based on a matrix-immune signature [J]. Discov Oncol, 2024, 15(1): 405. DOI: 10.1007/s12672-024-01265-w.

[20] LI E N, NI H J. Prognostic nomogram for early-stage cervical cancer in the elderly: a SEER database analysis [J]. Prev Med Rep, 2024, 41: 102700. DOI: 10.1016/j.pmedr.2024.102700.

[21] ZHOU W T, WANG Y L, LONG H M, *et al*. Construction of a nomogram prediction model for individualized prediction of cervical cancer risk: a retrospective study[J]. Int J Womens Health, 2025, 17: 4435-4446. DOI:10.2147/IJWH.S541374.

[22] DING H, KE Z H, XIAO X, *et al*. A predictive model using six genes DNA methylation markers to identify individuals with high risks of high-grade squamous intraepithelial lesions and cervical cancer [J]. Int J Womens Health, 2025, 17: 739-749. DOI:10.2147/IJWH.S494703.

(收稿:2025-12-10;修回:2026-02-08)

(编辑 张艺霖)