

High-Risk Clinical Characteristics and Risk Model for Unexpected Cervical Cancer Diagnosed After Surgery in Patients with LSIL/HSIL Biopsy

Guilong Pan^{1,*}, Wan Wang^{1,*}, Zhen Sun^{1,*}, Ran Chu¹⁻³, Jing Xue¹⁻³, Na Li¹⁻³

¹Department of Gynecology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, People's Republic of China; ²Shandong Provincial Health Commission Key Laboratory of Prevention and Treatment for Major Gynecological Diseases, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong Province, People's Republic of China; ³Jinan Key Laboratory of Diagnosis and Treatment of Major Gynecological Disease, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong Province, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jing Xue; Na Li, Department of Obstetrics and Gynecology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jingwu Road, Jinan, Shandong Province, People's Republic of China, Tel +86 18678890133; +86 18753157305, Email 18678890133@126.com; 18753157305@163.com

Background: Women with colposcopy-directed biopsy results of low-grade or high-grade squamous intraepithelial lesion (LSIL/HSIL) who are subsequently found to have cervical cancer after surgery likely represent missed preoperative malignancy. This study aims to identify risk factors associated with these missed diagnoses and to develop a preoperative risk prediction model.

Methods: Patients with LSIL/HSIL on colposcopy-directed biopsy who underwent cervical conization or hysterectomy at a single center were retrospectively included. Postoperative cervical cancer was the adverse outcome. Baseline characteristics were compared using chi-square tests. Predictors were identified by univariate and multivariable logistic regression to build a logistic regression model. Internal validation applied machine learning with five-fold cross-validation. Performance was evaluated by receiver operating characteristic (ROC) curve (AUC) and accuracy, with discrimination, calibration, and clinical utility assessed using ROC curves, calibration plots, and decision curve analysis (DCA).

Results: A total of 1271 patients were included, and 79 were diagnosed with cervical cancer after surgery. Risk factors significantly associated with postoperative cervical cancer diagnosis included age (OR, 1.0; 95% CI, 1.0–1.0, $p=0.045$), gravidity (OR, 1.2; 95% CI, 1.1–1.3, $p<0.001$), human papillomavirus 16 or 18 infection (OR, 2.6; 95% CI, 1.4–4.6, $p=0.002$), colposcope biopsy site numbers ≥ 4 (OR, 0.2; 95% CI, 0.1–0.3, $p<0.001$), ThinPrep Cytologic Test indicates high-grade squamous intraepithelial lesion (OR, 3.1; 95% CI, 1.1–8.5, $p=0.028$), and neutrophil-to-lymphocyte ratio (OR, 1.1; 95% CI, 1.0–1.3, $p=0.015$). In five-fold cross-validation, the machine learning algorithms yielded an AUC of 0.72–0.81 and an accuracy of 0.93–0.94 for the model. ROC curves, calibration curves, and DCA analysis demonstrate the model's excellent predictive capability and performance.

Conclusion: This study developed and internally validated a clinical prediction model to estimate the preoperative risk of occult cervical cancer in patients with LSIL/HSIL on colposcopy-directed biopsy. This model may support risk-stratified surgical decision-making and help reduce missed diagnoses.

Plain Language Summary: In clinical practice, a proportion of people with cervical cancer are still missed by colposcopy-guided biopsy. Some individuals receive a colposcopy-guided biopsy diagnosis of low-grade squamous intraepithelial lesion (LSIL) or high-grade squamous intraepithelial lesion (HSIL), but are subsequently found to have invasive cervical cancer only after cervical conization or hysterectomy. Such missed diagnoses may delay appropriate treatment and adversely affect prognosis.

This study retrospectively analyzed 1271 patients whose colposcopy-guided biopsies suggested LSIL or HSIL and who later underwent loop electrosurgical excision procedure, cold knife conization, or hysterectomy at a tertiary center. Among these patients, 79 were diagnosed with cervical cancer. Baseline clinical characteristics were compared between those with and without cervical cancer, and logistic regression analysis was used to identify high-risk clinical factors and to construct a nomogram-based risk assessment model.

Older age, higher gravidity, human papillomavirus infection, abnormal ThinPrep Cytologic Test results, increased neutrophil-to-lymphocyte ratio, and the number of colposcopic biopsy sites were associated with a higher likelihood of postoperative diagnosis of cervical cancer. These factors were integrated into a clinical prediction model that showed good ability to distinguish between individuals with and without occult cervical cancer, with favourable calibration and potential clinical utility.

These findings suggest that colposcopy and biopsy result alone may be insufficient to reliably exclude early cervical cancer. Applying this risk prediction model alongside routine clinical and laboratory information may help clinicians to identify individuals at higher risk of occult cervical cancer, optimize preoperative surgical planning, reduce missed diagnoses, and improve management of cervical lesions.

Keywords: colposcopy, missed diagnosis, cervical cancer, predictive model

Introduction

Cervical cancer remains one of the most prevalent malignant tumors among women worldwide, and its early detection and timely treatment are vital for improving patient prognosis.¹ Colposcopic biopsy, as a cornerstone of cervical cancer screening and early diagnosis, plays a pivotal role in evaluating cervical lesions.² However, despite its widespread use, the accuracy of colposcopic diagnosis is not absolute. Recent studies report that the diagnostic accuracy of pre-surgical colposcopy is approximately 80%, highlighting the influence of additional factors beyond the colposcopic procedure itself.^{3–6} These limitations pose significant challenges for clinical decision-making and treatment planning, particularly when discrepancies arise between biopsy results and subsequent surgical pathology.

Notably, even when colposcopic biopsy suggests precancerous lesions, a subset of patients is later diagnosed with invasive cervical cancer following cervical conization or total hysterectomy. In different countries and regions, the diagnostic missed rate of preoperative colposcopy is approximately 10%–20%.^{7,8} This diagnostic inconsistency not only complicates treatment planning but also raises critical concerns about the reliability and clinical applicability of colposcopic biopsy. The presence of missed diagnoses underscores the risk of treatment delays or insufficient intervention, which may adversely affect patient outcomes.

Underdiagnosis prior to surgery can lead to an underestimation of the severity of cervical lesions, potentially resulting in incomplete excision of cancerous tissue during conization.⁹ This increases the risk of recurrence and metastasis. These issues emphasize the urgent need for improved diagnostic strategies capable of enhancing the accuracy of cervical cancer detection and reducing the likelihood of missed diagnoses.^{10,11}

To address these challenges, this study aims to systematically investigate the clinical risk factors in patients whose colposcopic biopsy results suggest precancerous cervical lesions but who are ultimately diagnosed with cervical cancer following conization or total hysterectomy. By identifying these risk factors, we aim to develop a predictive model that integrates a comprehensive range of clinical variables to improve the diagnostic accuracy of colposcopic biopsy. Ultimately, this study seeks to address the limitations of current diagnostic approaches, minimize the rates of underdiagnosis, support more precise clinical decision-making, and enhance treatment outcomes for patients with cervical cancer.

Methods

Study Design and Patient Selection

This retrospective study collected data from Shandong Provincial Hospital, affiliated with Shandong First Medical University, during the period from 2009 to 2024. The inclusion criteria for patients were: (a) preoperative colposcopic pathology diagnosis of low-grade squamous intraepithelial lesion (LSIL) or high-grade squamous intraepithelial lesion (HSIL); (b) procedures performed include loop electrosurgical excision procedure (LEEP), cold knife conization (CKC) or hysterectomy. Exclusion criteria were: (a) the presence of other primary malignant tumors; (b) incomplete medical records or excessive missing data; (c) other specialized treatment before surgery. We divided all patients into two groups based on whether they were unexpectedly diagnosed with cervical cancer after surgery.

Baseline Characteristics

Clinical data collected included age, gravidity, parity, body mass index (BMI), history of conization, human papillomavirus (HPV) infection, and ThinPrep Cytologic Test (TCT) results—such as LSIL, HSIL, atypical squamous cells, cannot exclude HSIL (ASC-H), number of colposcopic biopsy sites, and atypical squamous cells of undetermined significance (ASC-US). Other variables included diagnostic and pathologic results of colposcopy, neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR) and other laboratory data. Surgical procedures performed included LEEP, CKC, and hysterectomy.

Endpoints

This study focused on unexpected postoperative diagnoses of cervical cancer as the clinical outcome. Specifically, this involved patients who underwent conization or hysterectomy, where postoperative pathology results indicated cervical cancer. Notably, all preoperative evaluations in these patients showed no evidence of potential malignancy.

Statistical Analysis

The clinical characteristics of patients were compared between the cancer group and the non-cancer group. Continuous variables (age, NLR) were modeled as continuous predictors in the logistic regression without categorization to preserve statistical power and avoid arbitrary cutoffs. Continuous variables were compared using the non-parametric Wilcoxon rank-sum test, with results presented as median and interquartile range. Categorical variables were compared using the chi-square test, with results expressed as *n* (%). A *p*-value of < 0.05 was considered statistically significant.

Logistic regression analysis was employed to identify clinical characteristics associated with unexpected postoperative diagnoses of cervical cancer. Continuous variables were modeled as continuous predictors in the logistic regression without categorization to preserve statistical power and avoid arbitrary cutoffs. Variables with a *p*-value < 0.05 from the univariate logistic regression analysis were included in the multivariate logistic regression analysis. Results were expressed as odds ratios (ORs), 95% confidence intervals (CIs), and *p*-values. Subgroup analyses of surgical procedures (LEEP/CKC group and hysterectomy group) assessed model discrimination, calibration, and clinical utility through the area under the receiver operating characteristic curve (AUC-ROC), calibration plots, and decision curve analysis (DCA). Based on the high-risk factors identified through logistic regression analysis, a nomogram was developed to predict the risk of unexpected postoperative diagnoses of cervical cancer. An individual risk score was calculated for each patient using the nomogram model. Patients were then classified into low- and high-risk groups using the cohort median score as the cutoff, and the proportion of unexpectedly diagnosed cervical cancer after surgery was compared between the two groups.

Two machine learning models—logistic regression and random forest—were subjected to internal cross-validation. To assess model performance and generalizability, a repeated 5-fold cross-validation strategy was employed, with 10 repetitions. Performance metrics included AUC-ROC for discriminative ability and accuracy for overall classification performance. Values are reported as means and standard deviations (SD).

Results

General Patient Characteristics

The study flowchart is presented in [Figure 1](#). In total, 1271 patients with LSIL or HSIL confirmed by colposcopy-guided biopsy were included, comprising 79 patients with an incidental postoperative diagnosis of cervical cancer and 1192 patients without cervical cancer. Baseline characteristics are summarized in [Table 1](#). The median age of the overall cohort was 40 years (interquartile range, 34–48). Colposcopic biopsy showed LSIL in 2.5% (32/1271) and HSIL in 97.5% (1239/1271) of patients. HPV16, HPV18, and HPV16/18 co-infection were detected in 55.6% (707/1271), 2.8% (35/1271), and 2.5% (32/1271) of patients, respectively.

Univariable and Multivariable Logistic Regression Analysis

Univariate and multivariate logistic regression analyses were performed to identify factors associated with an incidental postoperative diagnosis of cervical cancer. In univariate analysis, age (OR, 1.0; 95% CI, 1.0–1.0; *p* = 0.038), gravidity (OR,

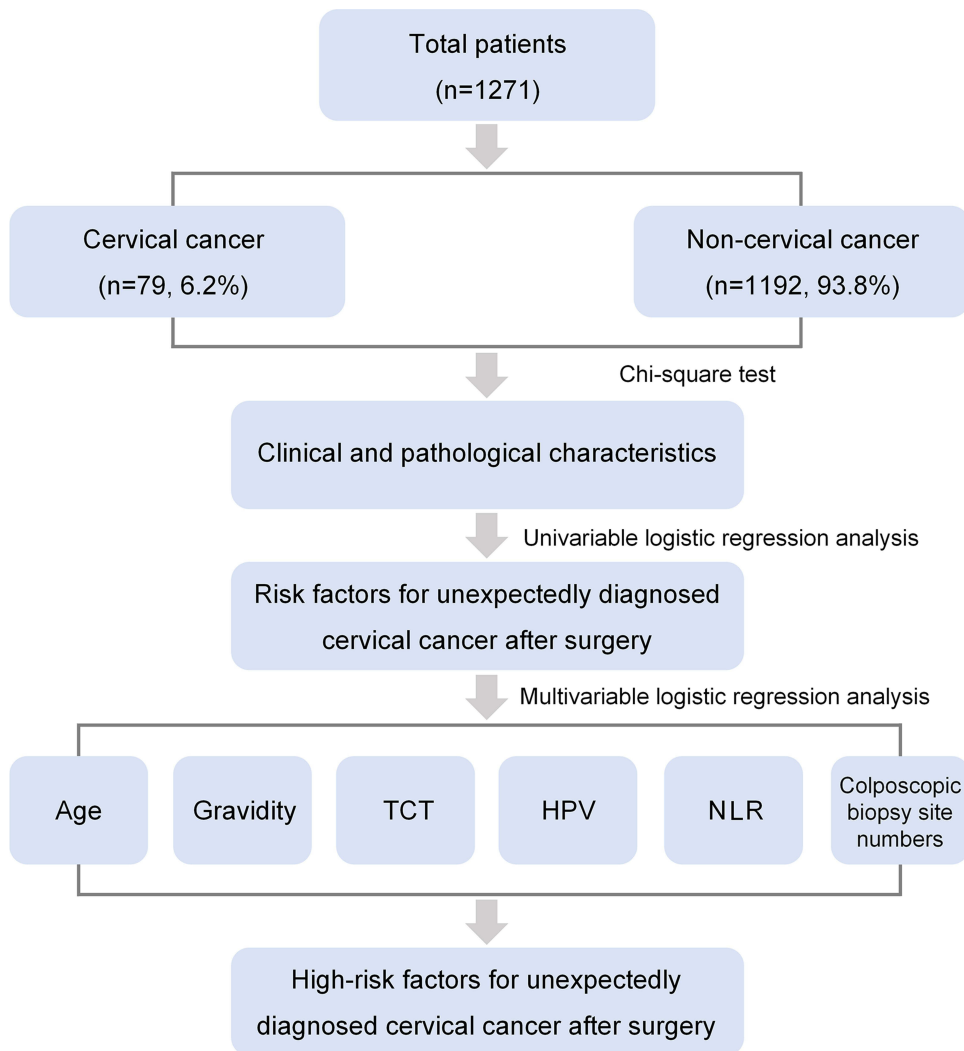


Figure 1 Study flowchart.

Abbreviations: TCT, ThinPrep Cytologic Test; HPV, human papillomavirus; NLR, neutrophil-to-lymphocyte ratio.

1.2; 95% CI, 1.1–1.3; $p < 0.001$), number of colposcopic biopsy sites ($p < 0.001$), HPV infection ($p = 0.003$), TCT results ($p = 0.030$), and NLR (OR, 1.2; 95% CI, 1.1–1.3; $p = 0.002$) were significantly associated with cervical cancer. Variables significant in univariate analysis were entered into a multivariable logistic regression model. Age (OR, 1.0; 95% CI, 1.0–1.0, $p = 0.045$), gravity (OR, 1.2; 95% CI, 1.1–1.3, $p < 0.001$), HPV 16 or 18 infection (OR, 2.6; 95% CI, 1.4–4.6, $p = 0.002$), colposcope biopsy sites ≥ 4 (OR, 0.2; 95% CI, 0.1–0.3, $p < 0.001$), TCT indicates HISL (OR, 3.1; 95% CI, 1.1–8.5,

Table 1 Comparison of Clinical Characteristics Between Cancer and Non-Cancer Patients

Characteristics	Total (n=1271)	Cancer (n=79)	Non-Cancer (n=1192)	P-value
Age (years)	40 (34–48)	44 (37–51)	40 (34–48)	0.015
Gravity	0 (0–4)	4 (0–5)	0 (0–4)	<0.001
Parity	2 (1–2)	2 (1–2)	2 (1–2)	0.238
Endocervical curettage (vs no)				0.280
No	601 (47.3)	42 (53.2)	559 (46.9)	
Yes	670 (52.7)	37 (46.8)	633 (53.1)	

(Continued)

Table 1 (Continued).

Characteristics	Total (n=1271)	Cancer (n=79)	Non-Cancer (n=1192)	P-value
Menopause				0.713
No	1049 (82.5)	64 (81.0)	985 (82.6)	
Yes	222 (17.5)	15 (19.0)	207 (17.4)	
Colposcopic biopsy site numbers				<0.001
<4	538 (42.3)	46 (58.2)	492 (41.3)	
≥4	642 (50.5)	11 (13.9)	631 (52.9)	
Unknown	91 (7.2)	22 (27.8)	69 (5.8)	
Transformation zone of cervical				0.090
1	286 (22.5)	17 (21.5)	269 (22.6)	
2	281 (22.1)	9 (11.4)	272 (22.8)	
3	532 (41.9)	40 (50.6)	492 (41.3)	
Unknown	172 (13.5)	13 (16.5)	159 (13.3)	
BMI (kg/m²)				0.577
<24	557 (43.8)	35 (44.3)	552 (43.8)	
≥24	363 (28.6)	19 (24.1)	344 (28.9)	
Unknown	351 (28.6)	25 (31.6)	326 (27.3)	
History of conization				0.051
No	1225 (96.4)	73 (92.4)	1152 (96.6)	
Yes	46 (3.6)	6 (7.6)	40 (3.4)	
HPV infection				0.002
HPV 16 or 18	742 (58.4)	61 (77.2)	681 (57.1)	
HPV 16 and 18	32 (2.5)	2 (2.5)	30 (2.5)	
Others	497 (39.1)	16 (20.3)	481 (40.4)	
TCT results				0.017
Normal	143 (11.3)	5 (6.3)	138 (11.6)	
ASC-US/ASC-H	450 (35.4)	28 (35.4)	422 (35.4)	
LSIL	169 (13.3)	3 (3.8)	166 (13.9)	
HSIL	341 (26.8)	30 (38.0)	311 (26.1)	
Unknown	168 (13.2)	13 (16.5)	155 (13.0)	
Diagnosis of colposcopy				0.631
Normal	23 (1.8)	1 (1.3)	22 (1.8)	
LSIL	133 (10.5)	5 (6.3)	128 (10.7)	
HSIL	303 (23.8)	20 (25.3)	283 (23.7)	
Unknown	812 (63.9)	53 (67.1)	759 (63.7)	
Pathological results of colposcopy biopsy				0.463
LSIL	32 (2.5)	1 (1.3)	31 (2.6)	
HSIL	1239 (97.5)	78 (98.7)	1161 (97.4)	
Laboratory test result				
HGB (g/L)				
Hemoglobin (g/L)	129.0 (119.0–136.0)	127.0 (112.0–138.0)	129.0 (120.0–136.0)	0.721
White blood cell (10 ⁹ /L)	5.5 (4.7–6.5)	5.6 (4.6–6.7)	5.5 (4.7–6.5)	0.518
Neutrophil (10 ⁹ /L)	3.2 (2.5–4.1)	3.1 (2.5–4.3)	3.2 (2.5–4.0)	0.521
Lymphocyte (10 ⁹ /L)	1.7 (1.4–2.1)	1.8 (1.4–2.0)	1.7 (1.4–2.0)	0.565
Platelet (10 ⁹ /L)	253.0 (217.0–296.0)	244.0 (193.0–296.0)	253.0 (217.0–296.8)	0.199
NLR	1.8 (1.4–2.5)	1.9 (1.4–2.5)	1.8 (1.4–2.4)	0.614
PLR	147.4 (119.4–183.1)	147.4 (119.4–191.8)	146.8 (117.7–182.8)	0.988

(Continued)

Table 1 (Continued).

Characteristics	Total (n=1271)	Cancer (n=79)	Non-Cancer (n=1192)	P-value
Conization of cervix				<0.001
LEEP	1142 (89.9)	53 (67.1)	1089 (91.4)	
CKC	64 (5.0)	7 (8.9)	57 (4.8)	
Hysterectomy	65 (5.1)	19 (24.1)	46 (3.9)	

Note: Values are median (interquartile range) or n (%).

Abbreviations: BMI, body mass index; HPV, human papillomavirus; TCT, ThinPrep Cytologic Test; ASC-US, atypical squamous cells of undetermined significance; ASC-H, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; LEEP, loop electrosurgical excision procedure; CKC, cold knife conization.

p=0.028), and NLR (OR, 1.1; 95% CI, 1.0–1.3, p=0.015) remained independent risk factors. Overall, HSIL on TCT, HPV16/18 infection, and fewer colposcopic biopsy sites (<4) appeared to make relatively larger contributions to the risk of an incidental postoperative diagnosis of cervical cancer. Accordingly, the final model incorporated age, gravidity, HPV infection, TCT results, NLR, and the number of colposcopic biopsy sites. Results are summarized in [Table 2](#).

Table 2 Univariable and Multivariable Logistic Regression Analysis of Risk Factors Associated with the Cervical Cancer

Characteristics	Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	p value	OR (95% CI)	p value
Age (years)	1.0 (1.0–1.0)	0.038	1.0 (1.0–1.0)	0.045
Gravidity	1.2 (1.1–1.3)	<0.001	1.2 (1.1–1.3)	<0.001
Parity	1.2 (1.0–1.6)	0.087		
Menopause (vs no)	1.1 (0.6–2.0)	0.713		
Colposcopic biopsy site numbers		<0.001		<0.001
<4	Reference		Reference	
≥4	0.2 (0.1–0.4)	<0.001	0.2 (0.1–0.3)	<0.001
Unknown	5.4 (2.7–10.5)	<0.001	2.8 (1.5–5.1)	0.001
Endocervical curettage (vs no)	0.8 (0.5–1.2)	0.281		
Transformation zone of cervical		0.104		
1	Reference			
2	0.5 (0.2–1.2)	0.124		
3	1.3 (0.7–2.3)	0.400		
Unknown	1.3 (0.6–2.7)	0.500		
BMI (kg/m²)		0.579		
<24	Reference			
≥24	0.8 (0.5–1.5)	0.509		
Unknown	1.1 (0.7–1.9)	0.620		
History of conization (vs no)	2.4 (1.0–5.8)	0.058		
HPV infection		0.003		0.008
HPV 16 or 18	2.7 (1.5–4.7)	0.001	2.6 (1.4–4.6)	0.002
HPV 16 and 18	2.0 (0.4–9.1)	0.369	2.0 (0.4–10.2)	0.394
Others	Reference		Reference	
TCT results		0.030		0.034
Normal	Reference		Reference	
ASC-US/ASC-H	1.8 (0.7–4.8)	0.222	2.1 (0.8–5.7)	0.149
LSIL	0.5 (0.1–2.1)	0.347	0.6 (0.1–2.6)	0.493
HSIL	2.7 (1.0–7.0)	0.047	3.1 (1.1–8.5)	0.028
Unknown	2.3 (0.8–6.7)	0.119	2.4 (0.8–7.4)	0.114

(Continued)

Table 2 (Continued).

Characteristics	Univariable Analysis		Multivariable Analysis	
	OR (95% CI)	p value	OR (95% CI)	p value
Diagnosis of colposcopy		0.641		
Normal	Reference			
LSIL	0.9 (0.1–7.8)	0.892		
HSIL	1.6 (0.2–12.1)	0.674		
Unknown	1.5 (0.2–11.6)	0.677		
Pathological results of colposcopy biopsy		0.473		
LSIL	Reference			
HSIL	2.1 (0.3–11.6)	0.473		
Laboratory test result				
HGB (g/L)				
Hemoglobin (g/L)	1.0 (1.0–1.0)	0.466		
White blood cell ($10^9/L$)	1.1 (1.0–1.3)	0.144		
Neutrophil ($10^9/L$)	1.2 (1.0–1.3)	0.047		
Lymphocyte ($10^9/L$)	0.8 (0.5–1.3)	0.377		
Platelet ($10^9/L$)	1.0 (1.0–1.0)	0.171		
NLR	1.2 (1.1–1.3)	0.002	1.1 (1.0–1.3)	0.015
PLR	1.0 (1.0–1.0)	0.374		

Note: Values are median (interquartile range) or n (%).

Abbreviations: OR, Odds ratio; CI, confidence intervals. BMI, body mass index; HPV, human papillomavirus; ASC-US, atypical squamous cells of undetermined significance; ASC-H, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; LEEP, loop electrosurgical excision procedure; CKC, cold knife conization.

The Evaluation of the Performance of Models in This Study

Based on the different surgical methods, we divided the patients into three groups, all patients group, conization group, hysterectomy group. The ROC curves are shown in Figure 2A. Three models had AUC values of 0.830, 0.848 and 0.703, respectively, suggesting that three models effectively differentiated between cervical cancer and non-cervical cancer. Figure 2B display calibration curves illustrating the level of agreement between cervical cancer predicted by the three models and the actual outcomes, with mean absolute errors of 0.009, 0.008, 0.022, respectively. Figure 2C illustrates the DCA regarding the clinical effectiveness of the three models at various threshold probabilities. Three models' DCA consistently lie above the reference lines, suggesting their practical value in clinical decision-making.

We determined a clinical model for predicting the risk of cervical cancer based on age, gravidity, number of colposcopic biopsy sites, HPV infection, TCT results, NLR. Figure 3A depicts a nomogram that evaluates the probability of cervical cancer based on the total points. Consider a 70-year-old woman with HSIL on colposcopic pathology who underwent colposcopic biopsies at three sites, has a gravidity of five, HPV 16 positive, and an NLR of 20. Using the nomogram, the corresponding points are 40 for age, 45 for three biopsy sites, 20 for gravidity, 22.5 for HPV 16, 40 for HSIL, and 67 for NLR, yielding a total risk score of 234.5 points. Based on the "Cancer possibility" scale at the bottom of the nomogram, this total score corresponds to an approximately over 80% probability of an unexpected postoperative diagnosis of cervical cancer. The distribution of risk scores in the overall cohort is shown in Figure 3B, with the median value of 100.8 indicated by the central line and the range denoted by the minimum and maximum values. Using 100.8 as the cutoff, patients were stratified into a low-risk group (score \leq 100.8) and a high-risk group (score $>$ 100.8). The proportion of postoperative incidental cervical cancer was higher in the high-risk group (73, 11.5%) than in the low-risk group (6, 0.9%) (Figure 3C).

Finally, as shown in Table 3, machine learning algorithms (logistic regression and random forest) were applied to further evaluate the model's predictive performance. Across 10 repetitions, the model achieved accuracies of 0.93–0.94 and AUC values ranging from 0.72 to 0.81, indicating good and stable discriminative performance.

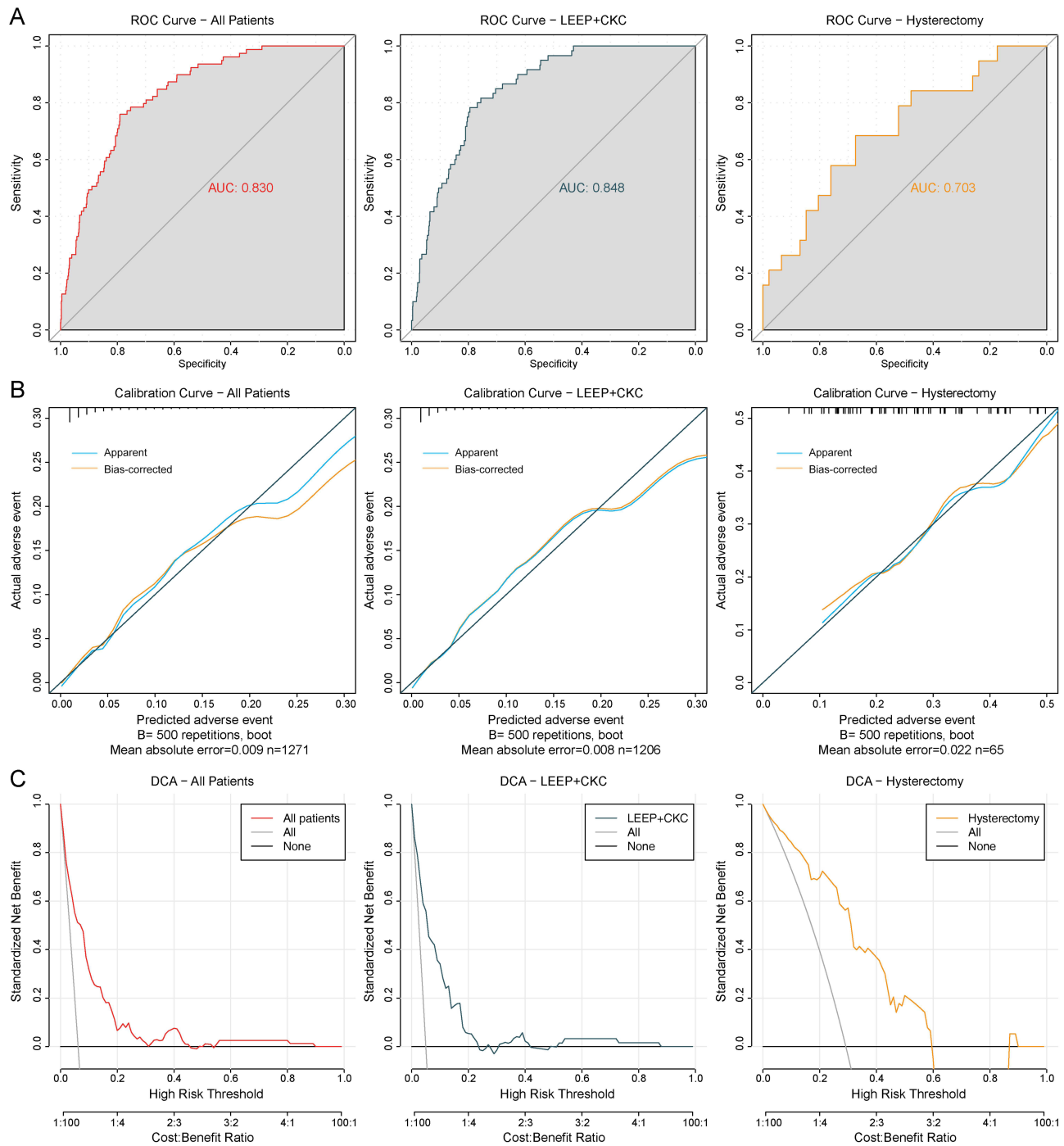


Figure 2 Receiver operating characteristic curve, calibration curve and decision curve analysis. **(A)** ROC curves for all patients, LEEP/CKC group, and hysterectomy group. **(B)** Calibration curves for all patients, LEEP/CKC group, and hysterectomy group. The dark diagonal line serves as the perfect benchmark; the blue curve represents apparent calibration, which may slightly overestimate performance; the orange curve represents bias-corrected calibration, providing a conservative estimate, and may show slight deviations. **(C)** DCA curves for all patients, LEEP/CKC group, and hysterectomy group.

Abbreviations: ROC, receiver operating characteristic; AUC, area under the receiver operating characteristic; LEEP, loop electrosurgical excision procedure; CKC, cold knife conization; DCA, decision curve analysis.

Discussion

A predictive model for the unexpected diagnosis of cervical cancer by postoperative pathology was developed and validated by a retrospective study of the clinical characteristics of 1271 patients with preoperative colposcopic pathology of precancerous cervical lesions. Notably, the vast majority of included patients had HSIL on colposcopic biopsy (97.5%), reflecting the clinical

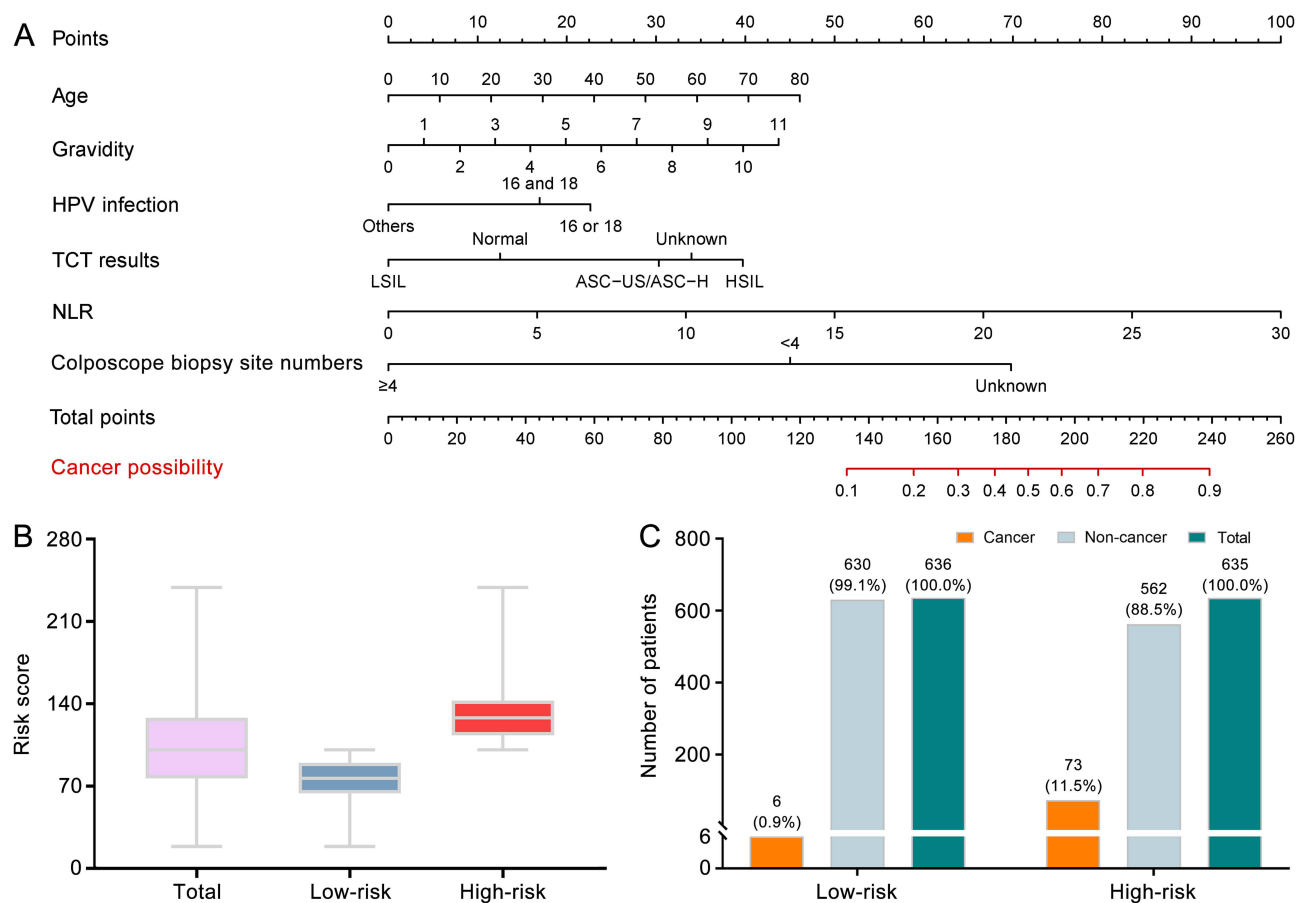


Figure 3 Nomogram model and risk stratification for postoperative incidental diagnosis of cervical cancer. **(A)** Nomogram for predicting the risk of incidental cervical cancer diagnosis after surgery. Points are assigned for each variable by drawing a line upward to the “Points” scale; the total score projected from the “Total points” axis corresponds to the predicted probability. **(B)** Distribution of risk scores across all patients, the middle line represents the median, and the upper and lower lines represent the maximum and minimum values, respectively. **(C)** Proportion of patients with postoperative incidental diagnosis of cervical cancer in the high-risk and low-risk groups. **Abbreviations:** HPV, human papillomavirus; TCT, ThinPrep Cytologic Test; ASC-US, atypical squamous cells of undetermined significance; ASC-H, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; NLR, neutrophil-to-lymphocyte ratio.

reality that most patients proceeding to conization/hysterectomy have high-grade lesions. Therefore, the developed model and factor weights are primarily applicable to patients with HSIL. This study showed that age, gravidity, HPV infection, TCT results, NLR, number of colposcopic biopsy sites were statistically associated with unexpected diagnosis of cervical cancer after surgery. Moreover, we chose five fold cross-validation to verify the predictive abilities of the models. The risk stratification derived from

Table 3 Model Performance Validation of Machine Learning Algorithm

Machine Learning Algorithm	Logistic Regression		Random Forest	
	AUC	Accuracy	AUC	Accuracy
Repeat 1	0.79	0.94	0.76	0.94
Repeat 2	0.79	0.94	0.75	0.94
Repeat 3	0.80	0.94	0.72	0.94
Repeat 4	0.80	0.94	0.74	0.94
Repeat 5	0.80	0.94	0.75	0.93
Repeat 6	0.81	0.94	0.76	0.93
Repeat 7	0.81	0.94	0.75	0.94
Repeat 8	0.81	0.94	0.75	0.94

(Continued)

Table 3 (Continued).

Machine Learning Algorithm	Logistic Regression		Random Forest	
	AUC	Accuracy	AUC	Accuracy
Repeat 9	0.81	0.94	0.75	0.94
Repeat 10	0.81	0.94	0.73	0.94
Mean	0.80	0.94	0.75	0.94
SD	0.01	0.00	0.01	0.00

Abbreviations: AUC, area under the receiver operating characteristic; SD, standard deviation.

our nomogram may help identify patients with a higher likelihood of occult cervical cancer and facilitate preoperative counseling and evaluation. Nevertheless, given the single-center retrospective design, the model should be used as an adjunct rather than a stand-alone determinant of surgical extent; individualized decisions should incorporate patient preferences and other clinical findings to avoid potential overtreatment.

We note that NLR has been associated with the unexpected diagnosis of cervical cancer after surgery. The association between inflammation and cancer was first noted in 1863, when it was first noted that inflammation associated with tumors plays an important role in tumor development.^{12,13} Early in tumor development, inflammation has a role in promoting cellular mutagenesis.¹⁴ A decline in lymphocytes predicts cellular immune injury, and increased numbers of neutrophils and platelets are considered responses to systemic inflammation.^{15–17} In the specific context of occult cervical cancer missed by biopsy, elevated NLR may reflect a tumor-promoting inflammatory microenvironment or subtle micro-invasion that is not adequately sampled during colposcopy. Systemic inflammation could facilitate early stromal invasion or angiogenesis even before macroscopic evidence is visible, thereby contributing to underdiagnosis at the preoperative stage.^{14,15} While NLR is nonspecific and influenced by multiple factors, its independent association in multivariable analysis supports its role as a readily available, low-cost preoperative biomarker in this setting. However, NLR may have some limitations due to the susceptibility of peripheral blood cell analysis results to factors such as circulatory capacity, infection, and nutritional status.

For patients with preoperative colposcopic pathology indicating precancerous cervical lesions and an unexpected postoperative finding of cervical cancer, part of the reason lies in the colposcopic biopsy collection process. Research has demonstrated that colposcopic accuracy is influenced by the colposcopists' level of subjective experience. Junior colposcopists have a tendency to overestimate findings in colposcopy, while senior colposcopists with more than five years of experience are more accurate in their assessments.^{18,19} The colposcopes determines the site and number of colposcopic biopsies, which determines whether or not the colposcopic biopsy pathology reflects the patient's true condition. It has been shown that a single biopsy from the most important area of the cervical transformation zone misses a large proportion of precancerous lesions, and that multiple targeted biopsies during colposcopy improve the detection of cervical cancer.^{20–23} In this study, the number of colposcopic biopsy sites was a predictor. Prior evidence shows that taking multiple biopsies improves the sensitivity and yield of colposcopy, highlighting biopsy number as a marker of sampling adequacy.^{20,21} At the same time, biopsy number may also reflect operator-related decision-making under diagnostic uncertainty (eg, experience and a lower threshold for sampling in less obvious or more extensive lesions), rather than patient biology alone.¹⁸ Therefore, this variable likely captures both examination/sampling adequacy and lesion characteristics, and its dual nature should be considered when applying the model clinically. Overall, given that colposcopic diagnostic performance varies with lesion visibility (eg, type 3 transformation zone) and is sensitive to methodological and quality-assurance factors, this predictor should be interpreted cautiously when translating the model into routine practice.^{24,25}

Surgery remains an effective treatment for early-stage cervical cancer. In patients with colposcopic pathology indicating precancerous lesions, integration of risk scoring based on independent factors may allow for appropriate intraoperative expansion of the surgical scope to excise potential tumor tissue, thereby reducing the likelihood of positive surgical margins. Furthermore, the clinical significance of hyperkeratotic cells in cervical cytology has not been fully elucidated. Hyperkeratosis constitutes a benign structural alteration in the cervical squamous epithelium that may obscure proliferative lesions and complicate accurate vaginal cytology.^{26,27} The indistinct demarcation between lesion tissue and

normal tissue further increases the risk of missed diagnosis; moreover, the small size of biopsy specimens and the limited extent of lesions contribute to a higher probability of diagnostic oversight.

In summary, the diagnostic accuracy of colposcopy requires further enhancement. Relying solely on colposcopic pathology is evidently inadequate for confirming cervical cancer diagnosis. The high-risk factors incorporated in this study represent well-established medical assessments that remain widely utilized in developing regions globally and are independent of the colposcopist's expertise level. Notably, AI-assisted colposcopic diagnosis has shown promising outcomes.²⁸ Emerging AI-assisted colposcopy systems have shown promising diagnostic performance and may provide objective, quantifiable imaging features (eg, transformation zone classification and lesion probability/extent measures) that could be integrated with routine clinical variables to enhance discrimination and calibration in future updates.²⁹ In parallel, computational biology and algorithmic approaches have been increasingly used to identify cervical cancer-related pathways/genes and to optimize feature selection, offering additional opportunities to incorporate molecular or image-derived biomarkers for refining individualized risk prediction.^{30,31} Therefore, when colposcopic pathology suggests cervical lesions, especially HSIL it is feasible to add age, gravidity, HPV infection, TCT results, Number of colposcope biopsy sites and NLR to consider cervical cancer progression comprehensively. This is used to determine the surgical approach, assist in reducing colposcopic leakage rates, and improve the prognosis of cervical cancer.

This study is a single Chinese center retrospective investigation with inherent limitations in patient selection, including potential selection bias due to its retrospective nature. Additionally, the time span of patients included in this study is relatively large, which may introduce temporal variations in diagnostic practices or data quality. The number of patients in the hysterectomy group is small, and we cannot exclude the possibility of a certain degree of overfitting in the AUC evaluation for this group. The lack of an external validation cohort further limits the model's generalizability. Future prospective, multicenter cohort studies in diverse populations are needed to externally validate and extend these findings.

Conclusion

In conclusion, we found that there is still room for improvement in the diagnosis of precancerous cervical lesions. Factors such as age, gravidity, HPV infection, Number of colposcope biopsy sites, TCT results and NLR affect the accuracy of cervical cancer diagnosis. Therefore, a comprehensive evaluation incorporating these factors should be performed in clinical practice to enhance diagnostic accuracy and reduce the possibility of secondary surgery.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author Na Li on reasonable request.

Ethical Approval

This retrospective study was approved by the Ethics Committee of Shandong Provincial Hospital (SWYX No. 2024-097). Informed consent was waived because patients had been discharged at the time of data collection. All data were anonymized and handled confidentially, in accordance with the Declaration of Helsinki.

Acknowledgments

We are grateful to all the patients who took part in this research.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by the Shandong Natural Science Foundation General Program (ZR2023MH326).

Disclosure

The authors report no conflicts of interest in this work.

References

- Xu M, Cao C, Wu P, Huang X, Ma D. Advances in cervical cancer: current insights and future directions. *Cancer Commun.* 2025;45(2):77–109. doi:10.1002/cac2.12629
- Duesing N, Schwarz J, Choschzick M, et al. Assessment of cervical intraepithelial neoplasia (CIN) with colposcopic biopsy and efficacy of loop electrosurgical excision procedure (LEEP). *Arch Gynecol Obstet.* 2012;286(6):1549–1554. doi:10.1007/s00404-012-2493-1
- İnal HA, Han O, Öztürk inal Z, Eren Karanis Mİ, Küçükosmanoğlu İ. Evaluation of concordance between loop electrosurgical excisional procedure and cervical colposcopic biopsy results. *J Turk Ger Gynecol Assoc.* 2024;25(1):13–17. doi:10.4274/jtgga.galenos.2023.2023-1-11
- Cantor SB, Cárdenas-Turanza M, Cox DD, et al. Accuracy of colposcopy in the diagnostic setting compared with the screening setting. *Obstet Gynecol.* 2008;111(1):7–14. doi:10.1097/01.AOG.0000295870.67752.b4
- Liu L, Liu J, Su Q, Chu Y, Xia H, Xu R. Performance of artificial intelligence for diagnosing cervical intraepithelial neoplasia and cervical cancer: a systematic review and meta-analysis. *EClinicalMedicine.* 2024;80:102992. doi:10.1016/j.eclinm.2024.102992
- Guo M, He M, Dang Y, et al. Predictors of para-aortic lymph node metastasis based on pathological diagnosis via surgical staging in patients with locally advanced cervical cancer: a multicenter study. *Cancer Lett.* 2025;616:217545. doi:10.1016/j.canlet.2025.217545
- Casarin J, Bogani G, Papadia A, et al. Preoperative minimization and risk of recurrence in patients undergoing laparoscopic radical hysterectomy for early stage cervical cancer: a multicenter study. *J Minim Invasive Gynecol.* 2021;28(1):117–123. doi:10.1016/j.jmig.2020.04.015
- Kamal M. Cervical pre-cancers: biopsy and immunohistochemistry. *Cytojournal.* 2022;19:38. doi:10.25259/CMAS_03_13_2021
- Jorgensen KA, Agusti N, Wu CF, et al. Fertility-sparing surgery vs standard surgery for early-stage cervical cancer: difference in 5-year life expectancy by tumor size. *Am J Obstet Gynecol.* 2024;230(6):663.e1–663.e13. doi:10.1016/j.ajog.2024.02.012
- Wenzel HHB, Schnack TH, Van der Aa MA, et al. Risk factors for lymph node metastasis in women with FIGO 2018 IA cervical cancer with a horizontal spread of >7mm. *Eur J Cancer.* 2024;212:115063. doi:10.1016/j.ejca.2024.115063
- Welp AM, Crawford M, O'Brien R, Sullivan SA, Duska LR. Presence of low volume metastases does not alter management in node-negative, early-stage cervical cancer patients who underwent postoperative adjuvant therapy: a retrospective cohort study. *Gynecol Oncol Rep.* 2023;51:101320. doi:10.1016/j.gore.2023.101320
- Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet.* 2001;357(9255):539–545. doi:10.1016/S0140-6736(00)04046-0
- Diakos CI, Charles KA, McMillan DC, Clarke SJ. Cancer-related inflammation and treatment effectiveness. *Lancet Oncol.* 2014;15(11):e493–e503. doi:10.1016/S1470-2045(14)70263-3
- Lu H, Ouyang W, Huang C. Inflammation, a key event in cancer development. *Mol Cancer Res.* 2006;4(4):221–233. doi:10.1158/1541-7786
- Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature.* 2008;454(7203):436–444. doi:10.1038/nature07205
- Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell.* 2010;140(6):883–899. doi:10.1016/j.cell.2010.01.025
- Ferrone C, Dranoff G. Dual roles for immunity in gastrointestinal cancers. *J Clin Oncol.* 2010;28(26):4045–4051. doi:10.1200/JCO.2010.27.9992
- Wei B, Zhang B, Xue P, et al. Improving colposcopic accuracy for cervical precancer detection: a retrospective multicenter study in China. *BMC Cancer.* 2022;22(1):388. doi:10.1186/s12885-022-09498-0
- Baum ME, Rader JS, Gibb RK, et al. Colposcopic accuracy of obstetrics and gynecology residents. *Gynecol Oncol.* 2006;103(3):966–970. doi:10.1016/j.ygyno.2006.06.002
- Gage JC, Hanson VW, Abbey K, et al. Number of cervical biopsies and sensitivity of colposcopy. *Obstet Gynecol.* 2006;108(2):264–272. doi:10.1097/01.AOG.0000220505.18525.85
- Pretorius RG, Belinson JL, Burchette RJ, et al. Regardless of skill, performing more biopsies increases the sensitivity of colposcopy. *J Low Genit Tract Dis.* 2011;15(3):180–188. doi:10.1097/LGT.0b013e3181fb4547
- van der Marel J, van Baars R, Rodriguez A, et al. The increased detection of cervical intraepithelial neoplasia when using a second biopsy at colposcopy. *Gynecol Oncol.* 2014;135(2):201–207. doi:10.1016/j.ygyno.2014.08.040
- Wentzensen N, Walker JL, Gold MA, et al. Multiple biopsies and detection of cervical cancer precursors at colposcopy. *J Clin Oncol.* 2015;33(1):83–89. doi:10.1200/JCO.2014.55.9948
- Li X, Zhao Y, Xiang F, et al. Evaluation of the diagnostic performance of colposcopy in the detection of cervical high-grade squamous intraepithelial lesions among women with transformation zone type 3. *BMC Cancer.* 2024;24(1):381. doi:10.1186/s12885-024-12156-2
- Brown HTJA, Tidy JA. The diagnostic accuracy of colposcopy - A review of research methodology and impact on the outcomes of quality assurance. *Eur J Obstet Gynecol Reprod Biol.* 2019;240:182–186.
- Fetissov F, Serres G, Arbeille B, et al. Argrophilic cells and ectocervical epithelium. *Int J Gynecol Pathol.* 1991;10(2):177–190. doi:10.1097/00004347-199104000-00006
- Xiao GQ, Emanuel PO. Cervical parakeratosis/hyperkeratosis as an important cause for false negative results of Pap smear and human papillomavirus test. *Aust N Z J Obstet Gynaecol.* 2009;49(3):302–306. doi:10.1111/j.1479-828X.2009.00998.x
- Kim S, Lee H, Lee S, et al. Role of Artificial Intelligence Interpretation of Colposcopic Images in Cervical Cancer Screening. *Healthcare.* 2022;10(3):468. doi:10.3390/healthcare10030468
- Takahashi T, Kobayashi Y, Sakurai R, et al. A systematic review of the application of Artificial Intelligence in colposcopy: diagnostic accuracy for cervical intraepithelial neoplasia and cervical cancer. *Clin Med Insights Oncol.* 2025;19:11795549251374908. doi:10.1177/11795549251374908
- Ali A, Mohan J, Nadaf TAA, Ravishankar H, Deepa KR. Bioinformatics-driven discovery of signaling pathways and genes influencing cervical cancer. *SN Computer Sci.* 2024;5(8):989. doi:10.1007/s42979-024-03347-6
- Ali A, Hulipal VR, Patil SS, et al. Consensus pattern selection from structured profile using multiobjective algorithm. *Int J Adv Sci Technol.* 2019;28(8):294–305.

International Journal of Women's Health

Publish your work in this journal

The International Journal of Women's Health is an international, peer-reviewed open-access journal publishing original research, reports, editorials, reviews and commentaries on all aspects of women's healthcare including gynecology, obstetrics, and breast cancer. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-womens-health-journal>

Dovepress
Taylor & Francis Group