Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prediction of Occult Invasive Cervical Cancer After Conization for High-Grade Squamous Intraepithelial Lesions: Model Development, Temporal Validation, and a Web-Based Tool

Jianmiao Chang
First Affiliated Hospital of Hebei North University
Xuntao Liu
First Affiliated Hospital of Hebei North University
Xiaojuan Liu
First Affiliated Hospital of Hebei North University
Wang Jing
First Affiliated Hospital of Hebei North University
Chunxing Ma
First Affiliated Hospital of Hebei North University
Jiangping Chen
jiangpingchen2006@163.com
First Affiliated Hospital of Hebei North University

Research Article

Keywords: HSIL, invasive carcinom conization machine-learning models web-based tool

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11374123/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Prediction of Occult Invasive Cervical Cancer After Conization for High-Grade Squamous Intraepithelial Lesions: Model Development, Temporal Validation, and a Web-Based Tool

Jianmiao Chang[1], Xuntao Liu[2], Xiaojuan Liu[1],Wang Jing[1], Chunxing Ma[1], Jiangping Chen[1*]

[1]Department of Gynecology, First Affiliated Hospital of Hebei North University, Zhangjiakou 075000, China

[2]Department of Clinical Laboratory, First Affiliated Hospital of Hebei North University, Zhangjiakou 075000, China

[*]**Corresponding authors**

Jiangping Chen, MD

12, Changqing Road, First Affiliated Hospital of Hebei North University

Zhangjiakou 075000, China

Email address: jiangpingchen2006@163.com

Abstract:

Objective: This study aimed to develop and temporally validate a prediction model for occult invasive carcinoma in patients with cervical high-grade squamous intraepithelial lesion (HSIL) undergoing conization and to develop a web-based prediction tool for individualized risk assessment.

Methods: We retrospectively included patients with cervical HSIL who underwent conization and were subsequently classified as having invasive carcinoma or non-upgraded disease according to the final histopathological findings. Least absolute shrinkage and selection operator (LASSO) regression were used for candidate predictor selection. Multivariable logistic regression and several machine-learning algorithms were developed and evaluated using internal cross-validation, followed by temporal validation in an independent cohort.The selected model was subsequently implemented as an R Shiny–based web prediction tool.

Results: Invasive carcinoma was detected after conization in approximately 6% of patients with cervical HSIL. LASSO selection identified age, high-grade cytology on TCT, HPV-16/18 positivity, involvement of 3–4 cervical quadrants, CIN2 or CIN3 on endocervical curettage (ECC), and CIN3 on biopsy as candidate predictors. In multivariable logistic regression, high-grade cytology, HPV-16/18 positivity, involvement of 3–4 quadrants, and CIN2 or CIN3 on ECC remained independently associated with histopathological upgrade. Among the evaluated algorithms, AdaBoost achieved the highest AUC in internal cross-validation and it was implemented as a web-based tool that provides individualized predicted probabilities and threshold-based risk classification using six routinely available preoperative variables.

Conclusion: Approximately 6% of patients with cervical HSIL in this cohort had occult invasive carcinoma identified after conization. Prediction models incorporating clinical and pathological variables showed good discriminative performance in the derivation cohort and retained satisfactory performance in temporal validation. A web-based prediction tool was subsequently developed to facilitate individualized risk assessment and may support clinical risk stratification for patients with cervical HSIL before conization.

Keywords: HSIL；invasive carcinom；conization；machine-learning models；web-based tool

Cervical cancer is the fourth most frequently diagnosed cancer among women worldwide. China bears a substantial share of this burden: in 2022, approximately 151,000 incident cases and 56,000 deaths were reported, accounting for 22.8% of global incident cases and 16.0% of global deaths, respectively[1]. The widespread implementation of cervical cancer screening has increased the detection of precancerous cervical lesions. High-grade squamous intraepithelial lesions (HSIL) are precancerous lesions that require appropriate management to prevent progression to invasive disease. Cervical conization by loop electrosurgical excision procedure (LEEP) or cold-knife conization (CKC) is the standard treatment for HSIL and provides both diagnostic and therapeutic information. Nevertheless, histopathological findings in conization specimens may be discordant those from preoperative colposcopy-directed biopsy, and occult invasive carcinoma may be detected after surgery. Reported concordance between conization histopathology and the preoperative diagnosis ranges from 64% to 85% in patients with cervical HSIL[2-4]. Recent studies have found invasive carcinoma in approximately 4% of conization specimens from these patients[5]. Because occult invasion remains difficult to predict before conization, treatment planning can be challenging, particularly for younger women who wish to preserve fertility. More extensive excision may compromise cervical function and increase the risks of preterm birth and miscarriage, whereas limited excision may leave residual disease or fail to detect invasive carcinoma. Preoperative risk assessment is also important for patients in whom conization is unsuitable because of unfavorable cervical anatomy.

Reliable preoperative identification of patients at high risk of histopathological upgrade to invasive carcinoma could support individualized treatment and reduce both excessive and insufficient intervention. We therefore retrospectively analyzed clinical data from patients who underwent cervical conization. Candidate predictors were selected by least absolute shrinkage and selection operator (LASSO) regression, and a multivariable logistic regression model was used as the baseline for comparison with several machine-learning algorithms. Performance was evaluated in an independent temporal validation cohort, and the selected model was subsequently implemented as a web-based tool. Our aim was to provide a practical method for estimating the risk of occult invasive carcinoma andsupporting individualized treatment planning.

1 Materials and Methods

1.1 Study Population

This retrospective cohort study used electronic medical records from patients treated at First Affiliated Hospital of Hebei North University between January 1, 2022, and December 31, 2024. Patients were eligible if cervical biopsy confirmed HSIL (CIN2 or CIN3), cervical conization was subsequently performed by LEEP or CKC at our hospital, and complete post-conization histopathological records were available. Patients missing any study variable were excluded. Additional exclusion criteria were adenocarcinoma in situ on cervical biopsy, suspected invasive disease in the biopsy report, pregnancy, and cervical conization during the preceding 3 years. The derivation cohort comprised 1,218 eligible patients: 76 (6.24%) had invasive carcinoma on post-

conization histopathology, and 1,142 showed no histopathological upgrading to invasive carcinoma. For temporal validation, we included a further 359 consecutive patients treated between January 1 and December 31, 2025; invasive carcinoma was identified in 22 patients (6.13%). The invasive carcinoma group included patients with HSIL on biopsy and invasive cervical carcinoma on post-conization histopathology, whether squamous cell carcinoma or adenocarcinoma. The non-upgrade group included patients whose biopsy showed HSIL but whose post-conization histopathology showed adenocarcinoma in situ, cervical HSIL, or a less severe lesion.

The following variables were collected: age, menopausal status, HPV genotype, multiple HPV infection, cytological findings (ThinPrep Cytologic Test [TCT]), preoperative biopsy pathology, the number of quadrants involved by HSIL, glandular involvement, and histopathological findings from endocervical curettage (ECC) and conization. HPV results were classified into three categories. The HPV-16/18 category comprised patients positive for either genotype. The HPV-52/58 category comprised patients positive for HPV-52 or HPV-58 but negative for both HPV-16 and HPV-18. The other-HPV category comprised patients with other high-risk HPV genotypes or a negative HPV test. Cytological findings were classified according to the Bethesda System[6]. The low-grade cytology category included negative for intraepithelial lesion or malignancy (NILM), atypical squamous cells of undetermined significance (ASC-US), and low-grade squamous intraepithelial lesion (LSIL). The high-grade cytology category included included atypical squamous cells, cannot exclude HSIL (ASC-H); HSIL; and atypical glandular cells (AGC). To record lesion extent on biopsy, the cervix was divided clockwise into four quadrants corresponding to clock-face positions 1-3, 4-6, 7-9, and 10-12, and the number of quadrants containing HSIL was recorded as a measure of lesion extent.

1.2 Statistical Analysis

All analyses were performed in R. Continuous variables were summarized as medians and interquartile ranges and compared between groups with the Mann-Whitney U test. Categorical variables were summarized as counts and percentages. Pearson's chi-square test was used when its assumptions for expected cell counts were met; otherwise, Fisher's exact test was applied. Candidate predictors for the multivariable logistic regression model were selected by LASSO. The machine-learning models comprised a decision tree, random forest, AdaBoost, gradient boosting decision tree (GBDT), XGBoost, and a backpropagation (BP) neural network. Hyperparameters were optimized by stratified repeated cross-validation within the derivation cohort. The temporal validation cohort was completely withheld from variable selection, model training, and hyperparameter tuning. Discrimination was assessed by the area under the receiver operating characteristic curve (AUC). Sensitivity, specificity, accuracy, positive predictive value, negative predictive value, and F1 score were used to describe classification performance, and the Brier score measured probabilistic prediction error. Agreement between predicted and observed event rates was examined with calibration curves. Decision-curve analysis was used to estimate net clinical benefit across threshold probabilities. $P < 0.05$ was considered statistically significant.

1.3 Ethics

The Institutional Ethics Committee of First Affiliated Hospital of Hebei North University approved the study (No. L2026246), and all procedures followed the principles of the Declaration of Helsinki. As this was a retrospective study involving the collection of clinical data, the Ethics Committee waived the requirement for informed consent.

2 Results

2.1 Baseline Characteristics

The derivation cohort included 1,218 patients, classified into the non-upgrade and invasive carcinoma groups. Median age was 44 years in the invasive carcinoma group and 39 years in the non-upgrade group. Menopause, high-grade cytology on TCT, HPV-16/18 positivity, involvement of 3-4 quadrants, CIN3 on ECC, and CIN3 on biopsy were more frequent in the invasive carcinoma group (all $P < 0.001$). The groups did not differ significantly in multiple HPV infection or glandular involvement (Table 1).

Table 1 Baseline characteristics and univariable comparisons in the derivation cohort

Variables	Total (n=1,218)	Non-upgrade group (n=1,142)	Invasive carcinoma group (n=76)	**P value**
Age, median (interquartile range)	39 (33, 47)	39 (33, 46)	44 (38, 54)	**<0.001**
Menopausal status, n (%)				**<0.001**
Yes	207 (17.00%)	182 (15.94%)	25 (32.89%)	
No	1011 (83.00%)	960 (84.06%)	51 (67.11%)	
TCT, n (%)				**<0.001**
high-grade	318 (26.11%)	277 (24.26%)	41 (53.95%)	
low-grade	900 (73.89%)	865 (75.74%)	35 (46.05%)	
HPV, n(%)				**<0.001**
16/18	641 (52.63%)	584 (51.14%)	57 (75.00%)	
52/58	335 (27.50%)	322 (28.20%)	13 (17.11%)	
Other genotypes	242 (19.87%)	236 (20.67%)	6 (7.89%)	
Multiple-type infection, n(%)				0.564
Yes	356 (29.23%)	336 (29.42%)	20 (26.32%)	
No	862 (70.77%)	806 (70.58%)	56 (73.68%)	
Lesion-involved quadrants, n (%)				**<0.001**
0–2	1017 (83.50%)	978 (85.64%)	39 (51.32%)	
3–4	201 (16.50%)	164 (14.36%)	37 (48.68%)	
ECC, n (%)				**<0.001**
CIN3	119 (9.77%)	72 (6.30%)	47 (61.84%)	
CIN2	117 (9.61%)	111 (9.72%)	6 (7.89%)	
Other	982 (80.62%)	959 (83.98%)	23 (30.26%)	
Glandular involvement, n (%)				0.344

Variables	Total (n=1,218)	Non-upgrade group (n=1,142)	Invasive carcinoma group (n=76)	P value
Yes	420 (34.48%)	390 (34.15%)	30 (39.47%)	
No	798 (65.52%)	752 (65.85%)	46 (60.53%)	
Biopsy result, n (%)				**<0.001**
CIN3	599 (49.18%)	533 (46.67%)	66 (86.84%)	
CIN2	619 (50.82%)	609 (53.33%)	10 (13.16%)	

2.2 LASSO Regression Analysis

Categorical predictors were represented by dummy variables before entry into the LASSO regression, and cross-validation was used to control model complexity. The coefficient paths showed how the estimate for each variable changed as the penalty increased. At λ.min, six predictor domains were retained: age, TCT, HPV status, lesion-involved quadrants, ECC findings, and biopsy results. Figure 1 shows the coefficient paths for all candidate predictors.

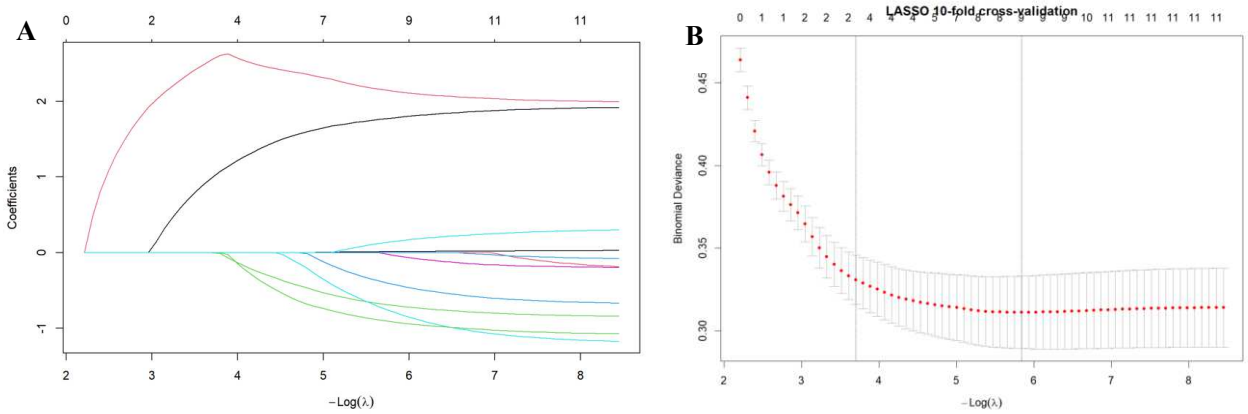


Figure 1.LASSO variable selection. (A) Coefficient paths for the dummy-coded predictors. The horizontal axis shows log(**λ**), and the vertical axis shows the regression coefficients. As lambda increases, the coefficients shrink toward zero. (B) Cross-validation curve. Red points show the mean binomial deviance at each **λ**, gray bars show one standard error, and the vertical dashed lines mark **λ**.min and **λ**.1se. The numbers above the plot indicate the number of nonzero coefficients.

2.3 Multivariable Logistic Baseline Model

The multivariable logistic regression baseline model included the six variables retained at λ.min: age, TCT findings, HPV genotype, the number of lesion-involved quadrants, ECC findings, and biopsy pathology. After adjustment for the other variables, high-grade cytology on TCT, HPV-16/18 positivity, involvement of 3-4 quadrants, and CIN2 or CIN3 on ECC were significantly associated with occult invasive carcinoma (Table 2).

Table 2 Multivariable logistic regression results

Variables	**OR**	**95% CI**	**P value**
Age (per 1-year increment)	1.022	0.994–1.050	0.129

Variables	OR	95% CI	P value
TCT: high-grade (reference: low-grade)	2.341	1.303–4.205	0.004
HPV: 52/58 (reference: other genotypes)	1.663	0.560–4.937	0.360
HPV: 16/18 (reference: other genotypes)	3.303	1.264–8.632	0.015
Lesion-involved quadrants: 3-4 (reference: 0-2)	6.742	3.492–13.015	<0.001
ECC: CIN2 (reference: others)	2.934	1.096–7.851	0.032
ECC: CIN3 (reference: others)	21.737	10.201–46.318	<0.001
Biopsy: CIN3 (reference: CIN2)	1.362	0.589–3.153	0.470

2.4 Development of Machine-Learning Models

The same six variables included in the multivariable logistic regression model were used to develop the machine-learning models. Within the derivation cohort, stratified cross-validation was used for model training and hyperparameter tuning, and performance in held-out folds was compared using the cross-validated AUC. Several ensemble methods achieved high AUC values, with AdaBoost achieving the highest value (0.888). AdaBoost was therefore selected on the basis of the derivation-cohort cross-validation results. The temporal validation cohort was completely withheld from model selection and played no role in model selection (Table 3and Figure 2).

Table 3 Cross-validated AUC values in the derivation cohort

Machine-learning models	Cross-validated AUC	Standard deviation (SD)
AdaBoost	0.888	0.054
GBDT	0.886	0.061
BP_NN	0.885	0.059
RF	0.885	0.052
XGBoost	0.879	0.047
Logistic	0.872	—
LASSO	0.860	—
DT	0.785	0.082

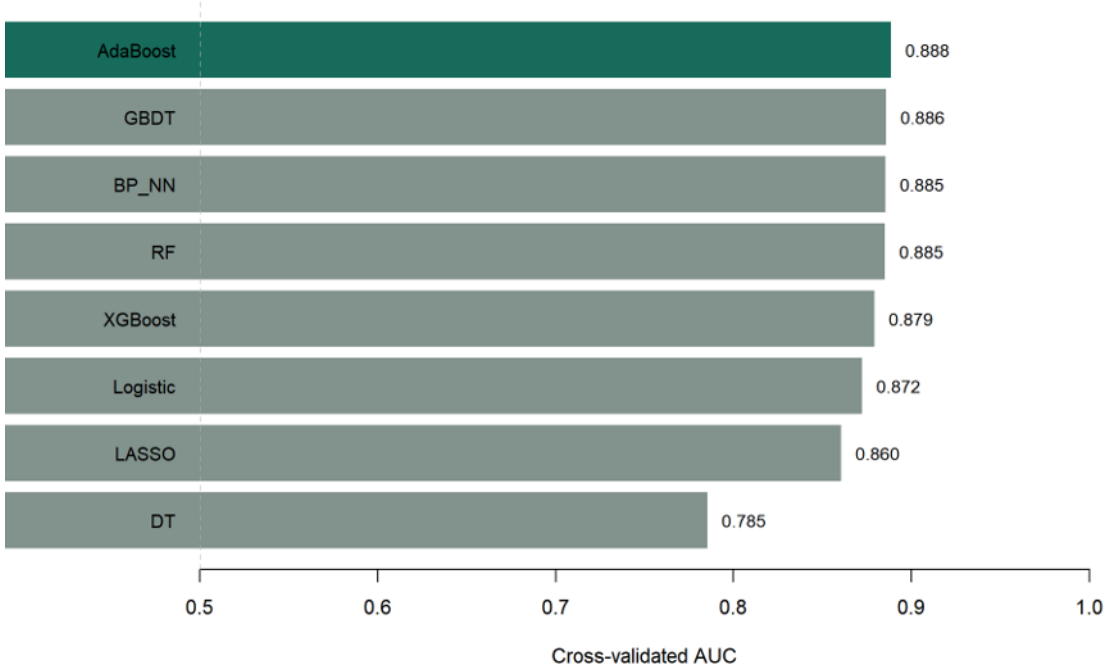


Figure 2. Cross-validated AUCs of the machine-learning models in the derivation cohort. The bars show AUC by model; the green bar identifies AdaBoost, the model selected in the internal comparison.

2.5 Temporal Validation

After training in the derivation cohort, each model was applied to the 2025 temporal validation cohort with its hyperparameters fixed. The classification threshold was also defined in the derivation cohort and was not re-optimized during validation. The derivation cohort contained 76 patients with invasive carcinoma (6.24%) and 1,142 in the non-upgrade group (93.76%); the corresponding numbers in the temporal validation cohort were 22 (6.13%) and 337 (93.87%). Thus, outcome prevalence was similar between cohorts, although the validation cohort contained fewer events. In temporal validation, AdaBoost achieved an AUC of 0.849 (95% CI, 0.755-0.943), sensitivity of 0.727, specificity of 0.760, and a Brier score of 0.106. Its performance was lower than the internal cross-validation result,although discrimination remained good.All models achieved an AUC >0.84 in the temporal validation cohort (Table 4). The ROC curves for LASSO, XGBoost, the BP neural network, and logistic regression were generally higher than that of the decision tree, which showed comparatively weaker discrimination (Figure 3). Because the validation cohort included few events, AUC point estimates should be interpreted together with their 95% CIs. Figure 4 displays the estimates and intervals to show their precision. Calibration curves comparing grouped predicted probabilities with observed event rates are presented in Figure 5. Decision-curve analysis was used to compare net clinical benefit across threshold probabilities, and the results are shown in Figure 6.The temporal validation data were used exclusively for performance evaluation and did not influence model selection or parameter tuning.

Table 4 Model performance in the temporal validation cohort

Machine-learning models	AUC	95%CI	Sensitivity	Specificity	Accuracy	F1 score	Brier score
BP_NN	0.941	0.904–0.978	0.864	0.887	0.886	0.481	0.039
LASSO	0.941	0.902–0.979	0.909	0.846	0.850	0.426	0.035
Logistic	0.937	0.897–0.977	0.909	0.837	0.841	0.412	0.035
XGBoost	0.931	0.891–0.971	0.864	0.840	0.841	0.400	0.050
RF	0.910	0.853–0.967	0.773	0.855	0.850	0.386	0.040
GBDT	0.872	0.772–0.973	0.864	0.754	0.760	0.306	0.039
AdaBoost	0.849	0.755–0.943	0.727	0.760	0.758	0.269	0.106
DT	0.841	0.743–0.939	0.727	0.938	0.925	0.542	0.036

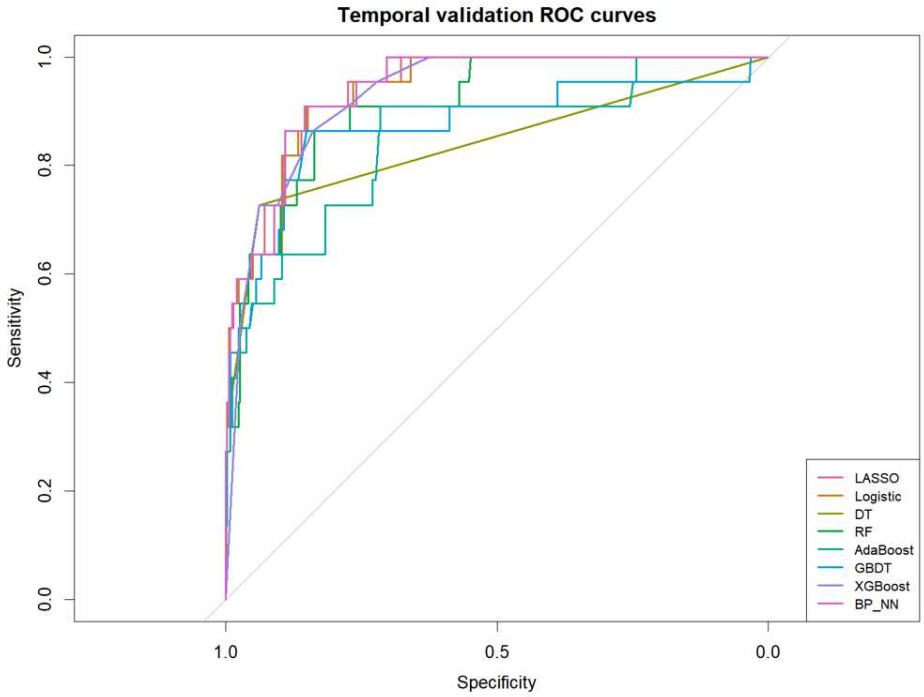


Figure 3. ROC curves in the temporal validation cohort. The horizontal axis shows 1 - specificity and the vertical axis shows sensitivity. Colors identify the prediction models, and curves closer to the upper-left corner indicate better discrimination. The diagonal line denotes random classification.

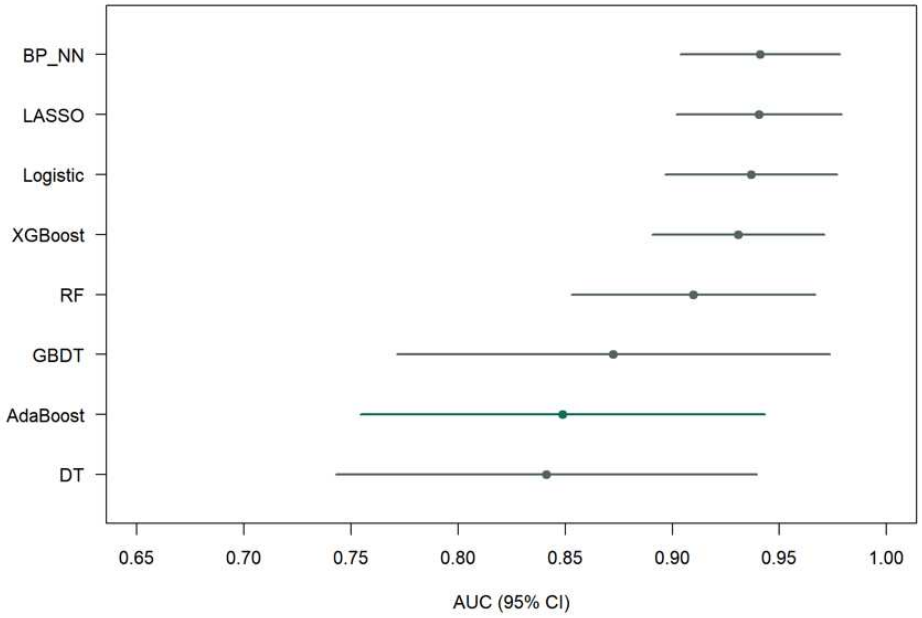
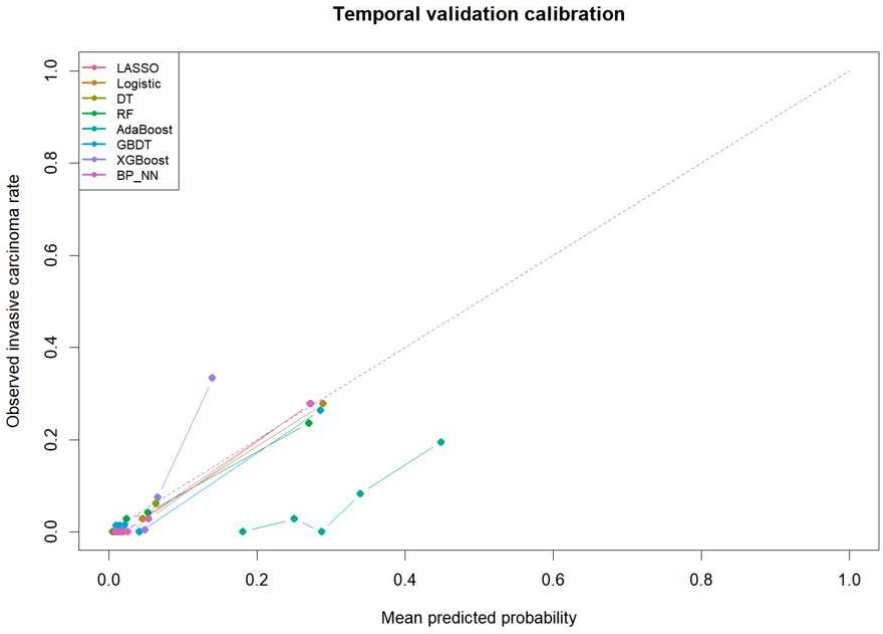


Figure 4. AUC estimates and 95% CIs for models evaluated in the temporal validation cohort. Points indicate the estimated AUCs, and horizontal lines indicate the 95% CIs. Wider intervals reflect the limited number of invasive carcinoma events in the validation cohort.



179

Figure 5. Calibration curves in the temporal validation cohort. The horizontal axis shows the mean predicted probability and the vertical axis shows the observed incidence of invasive carcinoma. The 45-degree line indicates perfect agreement; points closer to this line indicate better calibration.

180

181

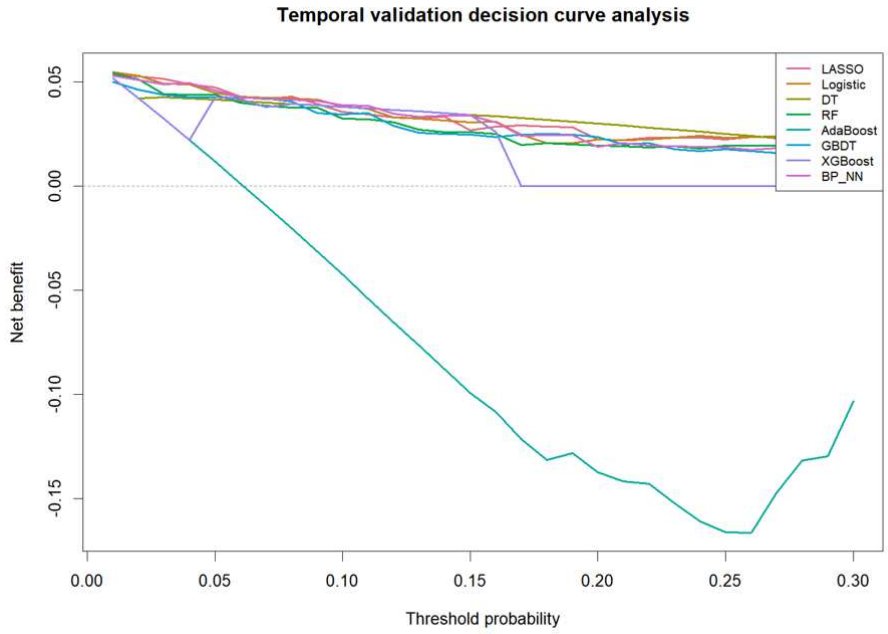


Figure 6 Decision curves **analysis** in the temporal validation cohort. The horizontal axis shows the threshold probability and the vertical axis shows net clinical benefit. Colors identify the prediction models. A model provides net benefit when its curve lies above the treat-all and treat-none reference strategies.

182

183 2.6 Web-Based Prediction Tool

184 The local R Shiny tool implemented the AdaBoost model, which achieved the highest cross-
185 validated AUC in the derivation cohort. After age, TCT findings, HPV genotype, number of lesion-
186 involved quadrants, ECC findings, and biopsy pathology are entered, the tool returns the predicted
187 probability of invasive carcinoma and a threshold-based risk classification. The classification
188 threshold was 35.0%, based on repeated cross-validated predictions in the derivation cohort. The
189 model uses a manageable set of clinically interpretable variables, allowing straightforward data
190 entry in routine practice. Its output is intended to support clinical risk assessment and does not
191 replace histopathological diagnosis.(https://gao462.shinyapps.io/hsil-risk-predictor/)

192 **3. Discussion**

193 3.1 Clinical Interpretation of Risk Factors

194 Univariable analysis and LASSO identified age, high-grade cytology on TCT, HPV-16/18
195 positivity, involvement of 3-4 quadrants, CIN2 or CIN3 on ECC, and CIN3 on biopsy as factors
196 associated with histopathological upgrade after conization. These findings accord with several
197 previous studies. Among the six variables entered into the multivariable logistic regression model,
198 patients with invasive carcinoma were older than those in the non-upgrade group, and menopausal
199 status also differed between groups. A recent study of 11,651 patients with high-grade cervical
200 intraepithelial neoplasia found that age independently predicted a final diagnosis of cervical cancer
201 after conization and was associated with occult invasive carcinoma[7]. Persistent high-risk HPV
202 infection, changes in immune clearance, and longer cumulative progression of cervical lesions may

contribute to this association. In our multivariable analysis, however, the adjusted effect of age was not statistically significant (OR=1.022, 95% CI: 0.994-1.050, P=0.129). Age may therefore reflect the patient's broader disease background rather than provide independent predictive information once TCT, HPV, ECC, and lesion extent are considered.

Preoperative TCT findings were associated with histopathological upgrade to invasive cervical carcinoma after conization. In the derivation cohort, high-grade cytology was present in 53.95% of patients with invasive carcinoma, compared with 24.26% of the non-upgrade group. The association remained significant after adjustment for other predictors (OR=2.341, 95% CI: 1.303-4.205, P=0.004). All participants had biopsy-confirmed HSIL before conization; within this population, the severity of the cytological abnormality may indicate not only disease but also an underlying risk of occult invasion. A 2025 study of patients with biopsy-confirmed high-grade cervical lesions likewise identified ASC-H and HSIL cytology as independent predictors of a final cervical cancer diagnosis[7]. This agreement supports the value of preoperative cytology in identifying patients at risk of occult cervical cancer.

HPV-16/18 positivity was another important predictor. In the derivation cohort, 75.00% of patients with invasive carcinoma were positive for HPV-16/18, compared with 51.14% of those without upgrade. After adjustment, patients positive for HPV-16/18 had higher odds of invasive carcinoma observed than those infected with other HPV genotypes (OR=3.303, 95% CI: 1.264-8.632, P=0.015), consistent with previous reports[2, 8]. In a study of 348 patients undergoing cervical conization, Shiga et al. found HPV-16 positivity to be an independent predictor of occult invasive cervical carcinoma (OR=5.26) [9]. Occult invasion should therefore remain a consideration in HPV-16/18-positive patients with HSIL even when preoperative biopsy has not identified carcinoma. HPV-52 and HPV-58 are also prevalent among Asian women. A large epidemiological survey in China reported prevalences of 3.5% and 2.1%, respectively[10]. In our cohort, however, HPV-52/58 was not independently associated with upgrade to invasive carcinoma and was more frequent in the non-upgrade group than in the invasive carcinoma group (28.20% vs. 17.11%). A relatively high prevalence of HPV-52/58 did not correspond to a higher observed risk of occult invasive carcinoma. Similarly, a study of HPV-positive women with normal cytology in northeastern China found no significant increase in histological HSIL+ with HPV-52 or HPV-58 infection[11]. Multiple HPV infection was also comparable between the non-upgrade and invasive carcinoma groups (29.42% vs. 26.32%, P=0.564), suggesting that detection of multiple HPV genotypes alone does not increase the likelihood of invasive carcinoma after conization for HSIL. One possible explanation is that the genotypes detected in a multiple infection do not all contribute to the same cervical lesion. Lesion-level molecular localization has shown that a lesion may be attributable predominantly to one genotype[12]. A large retrospective study also found no synergistic or cumulative oncogenic effect of multiple infection and suggested possible competition among genotypes[13]. Thus, PCR-based detection of multiple HPV types may describe the local presence of several genotypes rather than their joint contribution to lesion development and progression.

Involvement of 3-4 cervical quadrants was independently associated with occult cervical carcinoma (adjusted OR=6.742, 95% CI: 3.492-13.015). More extensive lesions may reflect more extensive HPV-associated high-grade epithelial disease, increasing the chance that a focal biopsy will miss the most severe area and underestimate disease before surgery. Previous studies have also reported more frequent residual disease after conization when more than two quadrants were involved[14]. In a recent study of 3,343 patients undergoing cold-knife conization for HSIL, involvement of at least three quadrants independently predicted positive resection margins (OR=3.264) [15]. These observations suggest that greater lesion extent may reflect both a higher disease burden and greater challenges in accurately assessing and treating the lesion. Patients with multiquadrant lesions therefore warrant particular attention for possible invasive carcinoma.

ECC was among the strongest predictors in this study. Relative to other ECC findings, CIN2 on ECC was associated with an approximately threefold increase in the risk of invasive carcinoma, whereas CIN3 was associated with an substantially higher odds (OR=21.737, 95% CI: 10.201-46.318, P<0.001). The latter was the largest effect estimate among the predictors among the predictors evaluated in the multivariable model. These findings suggest that ectocervical biopsy alone may underestimate disease in patients whose lesions extend into the endocervical canal. ECC samples this compartment and can therefore provide information not captured by a surface biopsy. A recent study of 11,651 patients with high-grade cervical lesions found that both ECC and cytology helped identify patients ultimately diagnosed with cervical cancer, and that a combined cytology-ECC indicator performed better than either measure alone[2]. The large OR for CIN3 on ECC in our study supports the inclusion of endocervical histopathology in risk-prediction models and underscores the role of curettage in detecting occult invasion within the canal. For patients with unfavorable cervical anatomy who cannot undergo conization, including postmenopausal women with pronounced cervical atrophy, repeat ECC may help establish a more definitive preoperative diagnosis.

Biopsy-confirmed CIN3 was significant in univariable analysis, but its adjusted OR decreased to 1.362 (95% CI: 0.589-3.151, P=0.470) in the multivariable model. This attenuation does not imply that biopsy pathology lacks clinical value. Rather, TCT, HPV, ECC, lesion extent, and biopsy pathology may contain overlapping information, reducing the independent contribution of biopsy after adjustment. Among patients with biopsy-confirmed HSIL, the risk of occult invasive carcinoma after conization should therefore not be assessed solely according to whether the biopsy shows CIN2 or CIN3; biopsy pathology is better interpreted alongside the other predictors in a multivariable model.

3.2 Comparison of Machine-Learning Models

We compared logistic regression, LASSO regression, a decision tree, random forest, AdaBoost, GBDT, XGBoost, and a BP neural network. In internal cross-validation, AdaBoost achieved the highest AUC, although the difference from GBDT, the BP neural network, and random forest were small. Conventional logistic regression estimates risk through an additive linear

relationship, whereas AdaBoost iteratively combines weak learners and can capture nonlinear relationships and interactions among predictors. This flexibility may be useful when clinical variables act jointly. However, the modest AUC differences observed in the derivation cohort do not establish that AdaBoost is uniformly superior to the other algorithms. Its stability requires confirmation in independent datasets.

Temporal validation was an important feature of this study. After training in the derivation cohort, model parameters and classification thresholds were locked before application to the 2025 cohort; the validation data were not used to reselect a model or optimize a threshold. AdaBoost achieved an AUC of 0.849, sensitivity of 0.727, and specificity of 0.760 in temporal validation, with some loss of performance relative to internal cross-validation. Larger temporal cohorts and multicenter external validation are needed before clinical implementation. The rate of histopathological upgrade was similar in the derivation and temporal cohorts (6.24% vs. 6.13%), and every model retained an AUC above 0.84 during temporal validation. These findings provide preliminary support of temporal transportability. Nevertheless, the validation cohort contained only 22 events, producing relatively wide confidence intervals and unstable calibration patterns. Larger temporal cohorts are needed to obtain more precise estimates.

3.3 Translational Value of the Web-Based Prediction Tool

Conventional nomograms require clinicians to calculate scores manually, which can hinder use in routine clinical settings. We therefore translated the selected model into an R Shiny tool deployed on shinyapps.io. Clinicians enter age, TCT findings, HPV genotype, lesion-involved quadrants, ECC findings, and biopsy pathology to obtain the predicted probability of invasive carcinoma and a threshold-based risk classification. The 35.0% threshold was derived from repeated cross-validated predictions in the derivation cohort and remained fixed during validation; it was not re-estimated using validation data. All six inputs are routinely available before conization and require no additional laboratory testing. Automated calculation also avoids errors from manual score reading and scale alignment, while providing immediate feedback. A public link allows access from computers and mobile devices, and the hosting arrangement can be adapted to future traffic and data-security requirements. These features may facilitate use of the model in routine practice. The tool is intended only to support clinical risk assessment and cannot replace histopathological diagnosis. Management of patients classified as high risk must still incorporate post-conization histopathology, and conservative treatment or follow-up in those classified as low risk should be selected cautiously.

3.4 Limitations

Several limitations should be considered. The retrospective design is susceptible to selection and information bias, and the completeness and measurement of recorded variables may have changed over time. Although temporal validation was performed, the validation cohort contained only 22 invasive carcinoma events. This limited event count may have reduced the stability of model estimates, particularly for sensitivity, calibration, and decision-curve analysis, and resulted in

relatively wide confidence intervals. The single-center setting also precluded geographic external validation, so multicenter studies are required to assess generalizability. In addition, the models included only clinical variables, including age, cytology, HPV status, lesion extent, ECC findings, and biopsy pathology, and did not incorporate the duration of persistent HPV infection, colposcopic images, or host DNA-methylation profiles. Host DNA-methylation markers have shown potential value for predicting high-grade cervical lesions and carcinogenesis[16, 17], and future models could integrate such biomarkers with clinical predictors and machine-learning methods. Finally, the web-based tool should be regarded as an adjunctive risk-assessment instrument rather than a diagnostic tool. Multicenter external validation, prospective clinical evaluation, and assessment of net benefit across risk thresholds are required before routine clinical use.

In this study, invasive carcinoma was detected after conization in approximately 6% of patients with cervical HSIL. Among the algorithms evaluated, AdaBoost performed showed the highest discriminative performance in the derivation cohort and retained good discrimination in temporal validation. The resulting model was implemented as a web-based prediction tool. The identified risk factors and prediction model may help recognize patients at high risk of occult invasive carcinoma, support clinical risk stratification and individualized management, and provide additional information for shared decision-making.

Ethics approval and consent to participate： The Institutional Ethics Committee of First Affiliated Hospital of Hebei North University approved the study (No. L2026246), and all procedures followed the principles of the Declaration of Helsinki. As this was a retrospective study involving the collection of clinical data, the Ethics Committee waived the requirement for informed consent.

Consent for publication： Not applicable.

Availability of data and materials： All data generated or analysed during this study are included in this published article.

Competing interests： The authors declare no conflicts of interest.

Funding: Zhangjiakou Science and Technology Research and Development Program Project (2021029D).

Authors' contributions: Conceptualization, CJM and CJP.; Software and data curation, LXT; Investigation, MCX，LXJ and WJ.; Writing—original draft preparation, CJM.; Writing—review and editing, CJP; Funding acquisition, CJM. All authors have read and agreed to the published version of the manuscript.

Acknowledgements: We sincerely thank all patients whose clinical records contributed to this study.

References

[1] Wu J, Jin Q, Zhang Y, et al. Global burden of cervical cancer: current estimates, temporal trend and future projections based on the GLOBOCAN 2022[J]. Journal of the National Cancer Center, 2025.

[2] Stuebs F A, Schulmeyer C E, Mehlhorn G, et al. Accuracy of colposcopy-directed biopsy in detecting early cervical neoplasia: a retrospective study[J]. Arch Gynecol Obstet, 2019,299(2):525-532.

[3] Jung Y, Lee A R, Lee S, et al. Clinical factors that affect diagnostic discrepancy between colposcopically directed biopsies and loop electrosurgical excision procedure conization of the uterine cervix[J]. Obstetrics & Gynecology Science, 2018,61(4):477.

[4] Giannella L, Mfuta K, Gardini G, et al. High-grade CIN on cervical biopsy and predictors of the subsequent cone histology results in women undergoing immediate conization[J]. European Journal of Obstetrics & Gynecology and Reproductive Biology, 2015,186:68-74.

[5] Zhang L, Tian P, Li B, et al. Risk-stratified management of cervical high-grade squamous intraepithelial lesion based on machine learning[J]. J Med Virol, 2024,96(10):e70016.

[6] Nayar R, Wilbur D C. The Pap test and Bethesda 2014[J]. Cancer Cytopathol, 2015,123(5):271-281.

[7] Yao X, Qie M, Kang L, et al. Assessing the risk of undetected cervical cancer in women with a biopsy of high-grade cervical intraepithelial neoplasia: predictive value of cytology and endocervical curettage (ECC)[J]. BMC Cancer, 2025,25(1).

[8] Liu Q, Yang J, Cheng H, et al. A Clinical Prediction Model for Pathologic Upgrade to Invasive Carcinoma Following Conization of Cervical High-Grade Squamous Intraepithelial Lesions[J]. Cancer Med, 2025,14(1):e70540.

[9] Shiga T, Taguchi A, Mori M, et al. Risk stratification of invasive cervical cancer diagnosed after cervical conization[J]. Jpn J Clin Oncol, 2023,53(12):1138-1143.

[10] Bao H, Jin C, Wang S, et al. Prevalence of cervicovaginal human papillomavirus infection and genotypes in the pre-vaccine era in China: A nationwide population-based study[J]. Journal of Infection, 2021,82(4):75-83.

[11] Zhang J, Zhang D, Yang Z, et al. The role of human papillomavirus genotyping for detecting high-grade intraepithelial neoplasia or cancer in HPV-positive women with normal cytology: a study from a hospital in northeastern China[J]. BMC Cancer, 2020,20(1).

[12] van der Marel J, Quint W G V, Schiffman M, et al. Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16[J]. International Journal of Cancer, 2012,131(6).

[13] Zhong F, Wang T, Li W, et al. Associations of Single Versus Multiple Human Papillomavirus Infections With the Prevalence of Cervical Intraepithelial Neoplasia 2/3 and Squamous Cell Carcinoma Lesions: Human Papillomavirus Type–Specific Attribution[J]. Laboratory Investigation, 2024,104(4):100328.

[14] Tasci T, Turan T, Ureyen I, et al. Is There Any Predictor for Residual Disease After Cervical

Conization With Positive Surgical Margins for HSIL or Microinvasive Cervical Cancer?[J]. Journal of Lower Genital Tract Disease, 2015,19(2):115-118.

[15] Zhang L, Zheng Y, Lei L, et al. Development of a machine learning-based model for predicting positive margins in high-grade squamous intraepithelial lesion (HSIL) treatment by Cold Knife Conization(CKC): a single-center retrospective study[J]. BMC Womens Health, 2024,24(1):332.

[16] Dick S, Heideman D A M, Berkhof J, et al. Clinical indications for host-cell DNA methylation markers in cervical screening and management of cervical intraepithelial neoplasia: A review[J]. Tumour Virus Research, 2025,19:200308.

[17] Bonde J, Floore A, Ejegod D, et al. Methylation markers FAM19A4 and miR124-2 as triage strategy for primary human papillomavirus screen positive women: A large European multicenter study[J]. International journal of cancer, 2021,148(2):396-405.