Development, validation, and clinical application of a machine learning model for risk stratification and management of cervical cancer screening based on full-genotyping hrHPV test (SMART-HPV): a modelling study

Binhua Dong,[a,b,s] Zhen Lu,[c,s] Tianjie Yang,[d,s] Junfeng Wang,[e] Yan Zhang,[f,a] Xunyuan Tuo,[g] Juntao Wang,[h] Shaomei Lin,[i] Hongning Cai,[j] Huan Cheng,[k] Xiaoli Cao,[l] Xinxin Huang,[m] Zheng Zheng,[d] Chong Miao,[n] Yue Wang,[a,b] Huifeng Xue,[o] Shuxia Xu,[p] Xianhua Liu,[p] Huachun Zou,[q,a,**] and Pengming Sun[a,b,r,*]

[a]Department of Gynecology, Fujian Key Laboratory of Women and Children's Critical Diseases Research, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, Fuzhou, Fujian, China
[b]Fujian Clinical Research Center for Gynecological Oncology, Fuzhou, Fujian, China
[c]School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong, China
[d]Department of Gynecology, Shenzhen Maternity & Child Healthcare Hospital, Shenzhen, Guangdong, China
[e]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands
[f]The State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, National Institute of Diagnostics and Vaccine Development in Infectious Diseases, School of Public Health, Xiamen University, Xiamen, Fujian, China
[g]Department of Gynecology, Gansu Provincial Maternity & Child Health-care Hospital, Lanzhou, Ganshu, China
[h]Department of Gynecology, Guiyang Maternal and Child Health Care Hospital, Guiyang, Guizhou, China
[i]Department of Gynecology, Shunde Women's and Children's Hospital of Guangdong Medical University, Foshan, Guangdong, China
[j]Department of Hubei Clinical Medical Research Center for Gynecologic Malignancy, Maternal and Child Health Hospital of Hubei Province (Women and Children's Hospital of Hubei Province), Wuhan, Hubei, China
[k]Department of Gynecology, Maternal and Child Health Hospital of Hongan County, Huanggang, Hubei, China
[l]Department of Gynecology, Maternal and Child Health Hospital of Gongan County, Jingzhou, Hubei, China
[m]The Ministry of Health, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, Fuzhou, Fujian, China
[n]Department of Information, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, Fuzhou, Fujian, China
[o]Center for Cervical Disease Diagnosis and Treatment, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, Fuzhou, Fujian, China
[p]Department of Pathology, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, Fuzhou, Fujian, China
[q]School of Public Health, Fudan University, Shanghai, China
[r]School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

Summary

Background **High-risk human papillomavirus (hrHPV) full genotyping facilitates risk stratification and efficiency in cervical cancer screening, widely verified and adopted in various screening settings. We aimed develop a cervical cancer predictive model that can guide referrals for colposcopy using hrHPV full genotyping data in a setting where screening rate is low.**

Methods **We developed, compared and validated four machine learning models (eXtreme gradient boosting [XGBoost], support vector machine [SVM], random forest [RF], and naïve bayes [NB]) for cervical cancer prediction, using data from a national cervical cancer screening project conducted in 267 healthcare centers in China. Cervical intraepithelial neoplasia grade 2 or worse (CIN2+) and CIN3+ were the primary and secondary outcomes. In various screening settings across China, the performance of discrimination was evaluated using area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, area under the precision–recall curve (AUPRC), and**

The Lancet Regional Health - Western Pacific 2025;55: 101480

Published Online xxx https://doi.org/10.1016/j.lanwpc.2025.101480

*Corresponding author. Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, 18 Daoshan Road, Fuzhou, Fujian 350001, PR China.
**Corresponding author. School of Public Health, Fudan University, 130 Dongan Road, Xuhui District, Shanghai, 200032, PR China.
E-mail addresses: fmsun1975@fjmu.edu.cn (P. Sun), zouhuachun@fudan.edu.cn (H. Zou).
[s]Co-first authors.

accuracy. Calibration and clinical utility were assessed with brier score, calibration curve and decision curve analysis (DCA).

Findings 1,112,846 women were recruited, of whom 599,043 were included in the analysis based on hrHPV full genotyping. Of these, 254,434 (age [years, median, IQR]: 48, 42–54), 297,479 (49, 43–55), 38,500 (37, 32–44), 1950 (38, 33–46), 1590 (53, 47–58), 779 (38, 31–49) and 4311 (40, 33–50) were in the development, temporal validation and external validation 1–5 datasets, respectively. The final simplified clinical risk prediction model includes hrHPV, number of HPV genotypes, cervical cytology, HPV16, HPV18, age, HPV52, HPV39 and gynecological examination. The final optimal XGBoost model for predicting CIN2+ showed good discrimination (AUROC, maximum 0.989 [0.987–0.992]; minimum 0.781 [0.74–0.819]), and calibration (brier score, maximum 0.118 [0.099–0.137]) in the five external validation sets. DCA showed that when the clinical decision threshold probability for optimal XGBoost model was less than 0.80, the model for predicting CIN2+ provided a superior standardized net benefit. The optimal XGBoost model obtained similar results in predicting CIN3+.

Interpretation We developed a cervical cancer screening risk prediction model that employs hrHPV full genotyping and simple test results to achieve risk prediction and stratified management for colposcopy referrals. This predictive tool is particularly suitable for settings with low screening rates.

Funding National Natural Science Foundation of China; Major Scientific Research Program for Young and Middle-aged Health Professionals of Fujian Province, China; Fujian Province Central Government-Guided Local Science and Technology Development Project; Fujian Province's Third Batch of Flexible Introduction of High-Level Medical Talent Teams; Fujian Provincial Natural Science Foundation of China; Fujian Provincial Science and Technology Innovation Joint Fund.

Keywords: Prediction model; Cervical cancer; Human papillomavirus; Full genotyping; Machine learning; China

Introduction

Cervical cancer is the fourth most common cause of death among women globally, with around 660,000 new cases and 350,000 deaths worldwide in 2022.[1] The incidence and mortality rates are still climbing, especially in low- and middle-income countries (LMICs), and there is a trend towards younger ages.[1,2] The main etiological factor for cervical cancer is the persistent infection with high-risk human papillomavirus (hrHPV).[3] Effective screening strategies can timely detect early-stage cervical cancer or precancerous lesions, enabling patients to receive early treatment and follow-up.[4] Studies have shown that introducing efficient cervical cancer screening programs in LMICs can result in a 34.2% reduction in cervical cancer mortality by around the year 2130.[5] Therefore, the implementation of high-efficiency cervical cancer screening strategies is crucial for quickly realizing the World Health Organization (WHO)'s objective of "Global Strategy to Accelerate the Elimination of Cervical Cancer".[6]

Low coverage of cervical cancer screening is a challenge faced by global health, currently about 67% of women aged 20–70 globally have never been screened for cervical cancer.[7] In particular, screening rates in LMICs are very low.[7] For example, the screening rate among women aged 35–49 in China is just 33%.[7] The main reasons for the low screening rates of cervical cancer are the limited availability of screening, the lack of accurate risk quantification and management, and an excessive number of cases referred for colposcopy examination, which adds burdens to the health system.[8]

Effective cervical cancer screening management techniques are the key to solving the problem of low screening coverage. The 2019 American Society for Colposcopy and Cervical Pathology (ASCCP) guidelines are the first to implement a risk-based, quantitative stratification and management approach for cervical cancer screening.[9] The 2019 ASCCP guidelines use a prevalence-incidence mixed prediction model.[9] The accurate prediction of Cervical intraepithelial neoplasia grade 3 and worse (CIN3+) incidence and prevalence depends on high-quality previous and current screening results.[9] In the development and external validation cohorts of the 2019 ASCCP model, the cervical cancer screening rate is 69% or higher, much higher than the 33% global screening rate among women aged 20–70.[7] Therefore, whether the ASCCP model is applicable in settings with low screening rates still needs further verification. Another obstacle is that in low screening rate settings like China,[7] many women lack previous screening results. However, previous and current screening results are important variables for the ASCCP model to achieve accurate prediction.

The hrHPV full genotyping test has been validated in cervical cancer screening across various settings, offering personalized risk stratification for screening

Research in context

Evidence before this study

Low coverage of cervical cancer screening is a challenge globally. We searched PubMed from its inception to July 2024 for studies using the keywords "machine learning", "risk estimation" and "cervical cancer screening" without any limitations on language or publication date. We found several studies, among which the most important one was the 2019 version ASCCP model study, the 2019 ASCCP guidelines for cervical cancer screening introduced the predictive model as tools to guide referrals for colposcopy for the first time. However, the data trained for the 2019 ASCCP model came from high-income settings with high screening rates. The model necessitates information about previous screening and are limited to some genotyping such as HPV16/18, which restricts its widespread use in low screening rate settings. The use of high-risk human papillomavirus (hrHPV) full genotyping in cervical cancer screening has facilitated risk stratification and greater cost-effectiveness, becoming widely utilized in settings with low screening rates. Several other studies have utilized machine learning to estimate cervical cancer risk, but their applicability in real-world cervical cancer screening has been limited due to small sample sizes, insufficient performance, poor interpretability, lack of hrHPV full genotyping, or the absence of representative data, hindering the practical implementation of these models.

Added value of this study

This study developed a cervical cancer prediction model for different settings with low screening rates, and it is a prediction model based on hrHPV full genotyping. This predictive tool employed hrHPV full genotyping and readily available test results to achieve excellent risk prediction and stratified management for colposcopy referrals. This study utilized the largest datasets in China to date. The strengths of our model include separate temporal validation and external validation cohorts from six distinct settings in China. In various external validations, our model demonstrates excellent performance in discrimination, calibration and interpretability, and might improve the effectiveness of real-world clinical practice through risk-based stratified management and interpretability. The model's online tool, SMART-HPV (available at http://cerviscreen.fjsfy.com/predict-tool/), could further facilitate its application across different settings globally.

Implications of all the available evidence

Our prediction model could simplify the management of cervical cancer screening, promote shared decision-making at the individual level, and facilitate risk-differentiated management at the health service level. This approach enables personalized colposcopy referrals for patients undergoing cervical cancer screening. Furthermore, to maximize the real-world clinical implementation of final models, we developed an online application. The accessibility and convenience of this tool could improve its clinical application in real-world clinical practice globally.

management.[10–13] With the advancements in hrHPV full genotyping technology, hrHPV full genotyping has been introduced into cervical cancer screening in multiple global settings, and it is now widely used in China.[11,13] The 2019 ASCCP prediction model performs genotyping on HPV16/18, but does not achieve full genotyping of hrHPV, which may lead to inaccurate risk quantification.[9] Developing a cervical cancer screening risk prediction model based on hrHPV full genotyping for risk stratification and management, is crucial for eliminating cervical cancer in settings with low screening rates.

Artificial Intelligence (AI) technology can extract key features from raw data for identification, and accurately predict the occurrence of malignant tumors.[14] This resolves the problems of low precision and limited adaptability that come with employing logistic regression.[15] In recent years, a few studies have explored the use of machine learning to perform high-throughput calculations on screening data to predict the risk of cervical cancer.[14,16] However, most of these studies only utilized data from a single center with a smaller sample size,[14,17] which lead to compromised predictive performance and limited clinical application. There have been no reports on cervical cancer risk prediction models incorporating full genotyping of hrHPV. Moreover, the absence of enough external validations and clinical utility assessments of these models limits their expandability for widespread application.

Our research aims to utilize national cervical cancer screening data from various settings to develop and validate an accurate and user-friendly cervical cancer risk prediction model. This model, based on hrHPV full genotyping data, aims to guide colposcopy referrals in clinical practice, enhancing the quality of cervical cancer screenings in settings with low screening rates.

Methods

Study design

This is a prediction model study for cervical cancer and precancerous lesions based on full genotyping of hrHPV, intended to guide clinical practices for colposcopy referrals. Our study consisted of five steps: development, calibration, validation, clinical utility assessment, and explanation. We first developed and compared machine

learning models with training data. Then the models were calibrated using another smaller dataset. Thereafter, internal validation, temporal validation with more recent data, and five external validations with data from multiple regions in China were performed. Next, we assessed clinical utility of the models. Finally, we evaluated the interpretability of models. To enhance the convenience of clinical practice with the model, we have also developed an online version of the final model. The study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.[18]

Study participants and cohorts

Our predictive model study conducted to analyze data from a national cervical cancer screening program carried out across 267 healthcare centers in China between August 2014 and December 2023. All data used for the model are based on cervical cancer screening data with hrHPV full genotyping, which must include the 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68 types of hrHPV recognized by the WHO. Women undergoing cervical cancer screening all utilize full genotyping of hrHPV test as an initial screening method. The biological sex of all participants in this study is female. Participant-level data for this study were gathered from six regions in China, with details as follows: Fujian Province (data from January 2014 to December 2018 and 2021 were treated as development cohort, data from 2020 to 2022 as temporal validation cohort), Shenzhen City (external cohort 1), Foshan City (external cohort 2), Hubei Province (external cohort 3), Guizhou Province (external cohort 4) and Gansu Province (external cohort 5). The selection of the external validation cohort considered different screening populations (community screening and hospital opportunistic screening), geographical locations, and income levels, aiming to verify the model's predictive performance in various environments across China. Detailed cohort information can be found in Appendix S1. We recruited women aged 25–64 with more than one year of sexual history, who are not pregnant, and have no history of cervical surgery as participants. All data were de-identified and anonymized, with no biological specimens collected. The detailed process for cervical cancer screening and diagnosis is provided in Appendix S1 and S2. This study was approved by the Ethics Committee of Fujian Provincial Maternal and Child Health Hospital (2023KY141). According to the CIOMS 2016 version of the International Ethical Guidelines for Health-Related Research Involving Humans,[19] our study is a retrospective data analysis, all the personal information such as names and phone numbers has been removed from the data of all participants, and it is no longer possible to contact the participants themselves. The additional risks to the participants caused by our study are less than the minimal risks (risks that can be obtained in normal life). Therefore, we submitted an application for exemption from informed consent to the ethics committee and got the approval. This study has been registered on ClinicalTrials (NCT06204133).

Outcomes

Cervical intraepithelial neoplasia 2 or worse (CIN2+) is defined as the primary outcome. CIN2+ is considered the treatment threshold in multiple cervical cancer screening guidelines, and it is used as the primary outcome measure in most studies.[20] Cervical intraepithelial neoplasia 3 or worse (CIN3+) is designated as secondary outcome, as women with CIN3 have a high rate of disease progression. Timely and effective treatment of women with CIN3+ can prevent progression to invasive cervical cancer. The diagnosis and treatment of women with CIN3+ is increasingly emphasized, which is why it was included as an outcome in this study.[20]

Candidate predictors

Based on the consensus of nine cervical cancer diagnostic experts (PS, HX, SX, BD, ZZ, HC, JW, SL, XT) and two statistical experts (HZ, ZL) from China, and one methodologist in prediction modelling (JW) from the Netherlands, with a systematic review of published guidelines on cervical cancer diagnosis,[9,21,22] the following candidate predictors were taken into consideration: demographic variables, history of cervical cancer screening, vulva and vaginal examination variables, pelvic examination variables, HPV genotyping testing, cervical cytology testing. Further details are listed in Appendix S1. We reviewed clinically relevant features collected, all with <2% missing data. We created missing data indicators for each feature, because this effectively creates a reflection on both observed and missing data patterns.[23] We included these missing data indicators into our modelling analyses.

Model development

The eXtreme gradient boosting (XGBoost) algorithm was employed to predict the outcomes of interest,[24,25] as per our previous work.[24] XGBoost utilizes iteratively decision trees to progressively enhance prediction performance and is widely applicable in classification tasks. Compared to traditional logistic regression, XGBoost demonstrates superior performance in handling high-dimensional sparse and non-linear data, and it can automatically handle missing values.[25] Another three machine learning algorithms were assessed to test the assumption that XGBoost would have the highest prediction performance: support vector machine (SVM), random forest (RF), and naïve bayes (NB).

From development cohort, we used stratified sampling to randomly select 80% individuals to train models and 10% individuals to calibrate the models. 10-fold cross-validation was used to tune hyperparameters, and optimal hyperparameters were acquired by Bayesian

optimization.[24,26] The remaining 10% individuals were used for internal validation. To address class imbalance, we experienced with three methods: oversampling, undersampling, and the use of class weights. We calibrated the models using an isotonic regression algorithm with tenfold cross-validation in order to obtain a risk probability more relevant to the true incidence probabilities of outcomes in this study.

To assess the contribution of each candidate factor to predicting the occurrence of CIN2+/CIN3+, the SHapley Additive exPlanations (SHAP) method is employed to evaluate the importance of all features.[27] Considering clinically availability and accessibility, integrate the judgments of clinical practice experts, pathology experts and statistical experts, and importance ranking of predictors are synthesized. Based on these, we selected key features for simplified prediction modelling.

Model evaluation

To assess discrimination of models, we used five evaluation metrics: area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, area under the precison–recall curve (AUPRC), and accuracy. We selected AUROC and AUPRC for performance assessment. Calibration performance was assessed by Brier score and calibration curve. We compared the prediction performances of the four machine learning algorithms on the internal validation, temporal validation, and external validation data.

Furthermore, decision curve analysis (DCA) was utilized to assess the clinical utility of the models. DCA considers the potential harm resulting from both true and false positives, integrating threshold probabilities and model standardized net benefit to present the decision curve in a concise graphical format.[17,24] The SHAP method support interpretable output from machine learning algorithms. With the measurement of local feature interaction, it could provide the global understanding of the models. We used SHAP algorithm to illustrate the models' explainability. Moreover, an online browser version of the final simplified models is also available for external use.

Statistical analysis

Age and the number of HPV types infected were described using the median (interquartile range). Categorical variables were described using counts and proportions. We used bootstrapping with 500 resamples to get 95% confidence intervals (CI) of the metrics. All statistical analyses and modeling were conducted using Python (version 3.11.6) and R (version 4.2.3).

Role of the funding source

The funders had no role in study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to submit for publication.

Results

Characteristics of cohorts

This study recruited 1,112,846 women participating in the national cervical cancer screening program in China. The detailed participant selection process was shown in Fig. 1. To summarize, 494,508 women were excluded for not undergoing hrHPV full genotyping, and 19,295 women were excluded due to incomplete screening results. Ultimately, 599,043 women were included in the analysis. The development, temporal and external validation cohorts included 254,434, 297,479, 38,500, 1950, 1590, 779, and 4311 women, respectively. The infection rates of hrHPV, HPV16, HPV18, HPV39, and HPV52 among women in the development cohort were 10.66%, 1.26%, 0.65%, 0.88%, and 3.54% respectively. The incidence rates for CIN2+ and CIN3+ were 0.73% and 0.70%, respectively. The demographic characteristics, gynecological examination, hrHPV genotyping, cervical cytology and cervical histopathology results for the development cohort, temporal validation cohort, and external validation cohorts are presented in Table 1 and Supplementary Tables S1 and S2.

Predictors importance

Based on assessment of predictor importance (Fig. 2 and Supplementary Table S3) for CIN2+/CIN3+, consensus of experts, previous studies and guidelines, the final panel of nine key predictors were hrHPV infection, number of genotypes with HPV infected, cervical cytology testing, positive for HPV16, positive for HPV18, age, positive for HPV52, positive for HPV39, and gynecological examination.

Discrimination and calibration of machine learning models

The overall predictive performance of the four models in the internal, temporal, and external validation datasets are depicted in Fig. 3 and Supplementary Tables S4 and S5. In the internal validation, the final simplified XGBoost model showed the best AUROC (0.991, 95% CI 0.989–0.993), sensitivity (0.994, 95% CI 0.981–1.0), specificity (0.955, 95% CI 0.953–0.958), AUPRC (0.426, 95% CI 0.359–0.493), accuracy (0.956, 95% CI 0.953–0.959) and brier score (0.005, 95% CI 0.004–0.006) for predicting CIN2+. In the temporal validation cohort, the XGBoost model also showed good discrimination (AUROC 0.985 [95% CI 0.984–0.987]) and calibration (brier score 0.005 [95% CI 0.004–0.005]) for predicting CIN2+. In the five external validation sets, the final simplified XGBoost model showed good discriminability in predicting CIN2+, the maximum AUROC, sensitivity, specificity, AUPRC and accuracy were 0.989 (95% CI 0.987–0.992), 0.856 (95% CI 0.819–0.885), 0.981 (95% CI 0.979–0.982), 0.55 (95% CI 0.501–0.597), 0.979 (95% CI 0.978–0.981), respectively, and the minimum AUROC, sensitivity, specificity, AUPRC and accuracy

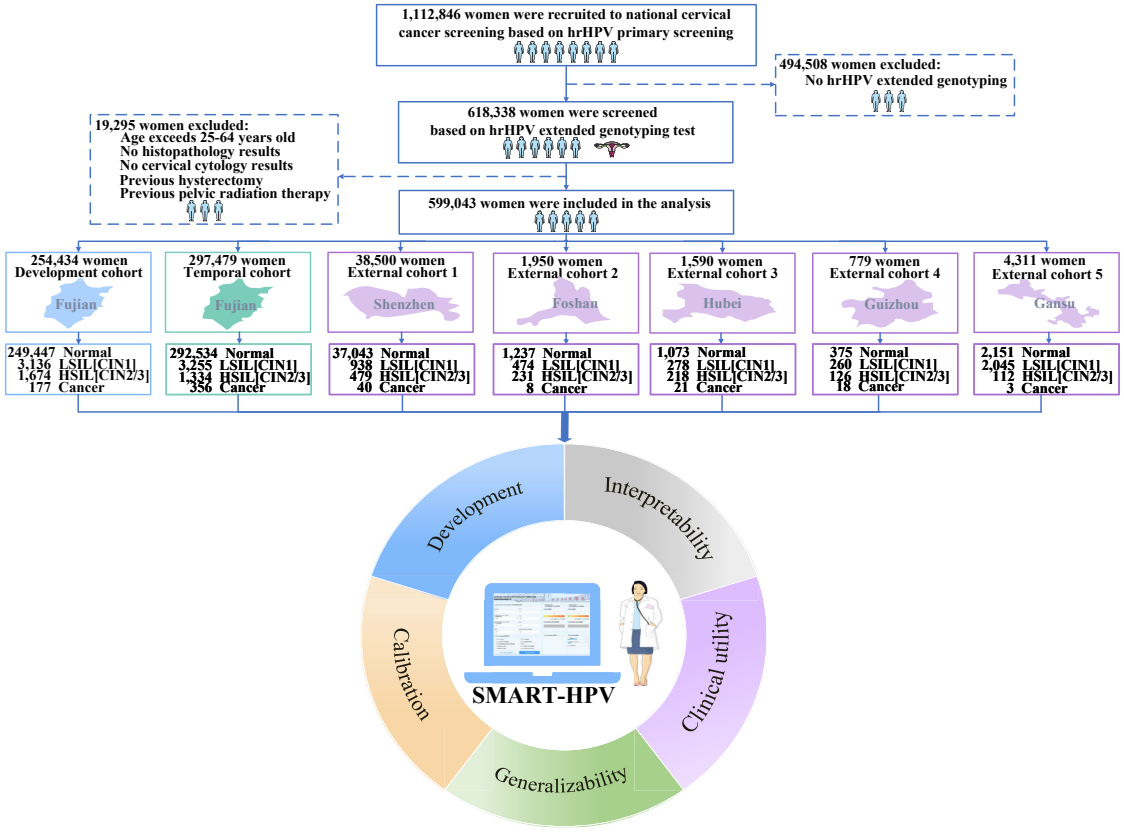

***Fig. 1:* Overall flowchart of the study**. hr-HPV, high-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesions; HSIL, high-grade squamous intraepithelial lesions; CIN, cervical intraepithelial neoplasia.

were 0.781 (95% CI 0.74–0.819), 0.959 (95% CI 0.917–0.989), 0.348 (95% CI 0.333–0.363), 0.146 (95% CI 0.086–0.207), 0.363 (95% CI 0.35–0.379), respectively. During internal and external validation, the Random Forest, Support Vector Machine and Naive Bayes models performed quite well in predicting CIN2+ outcomes, although not as well as the XGBoost model. Regarding the calibration performance of the four models, the calibration curves (Fig. 3) indicates that the XGBoost model has the best calibration in predicting CIN2+ outcomes, with the maximum and minimum values of the brier score were 0.118 [0.099–0.137] and 0.01 [0.009–0.011]) in the five external validation sets (Supplementary Table S6). In internal, temporal and external validations, the XGBoost model exhibited the best performance and calibration in predicting CIN3+ (Fig. 3 and Supplementary Table S6). Confusion matrixes were shown in Fig. 3e and f.

Clinical utility of models

Fig. 4a and b display net benefits of four models for CIN2+ and CIN3+ predictions, respectively, by comparison with strategies in which none or all of the women underwent colposcopy (i.e., treat none or treat all). Overall, XGBoost models had higher net benefits compared with the two reference strategies and the other models when the probability thresholds were below 0.61. Specifically, for CIN2+, using XGBoost model could yield net benefits for thresholds below 0.80, and it offered net benefits for thresholds lower than 0.61 for CIN3+.

Model explanation

Fig. 4c–f show the SHAP plots of CIN2+ and CIN3+ outcomes of four patients (with and without CIN2+/CIN3+). The first patient was a woman aged 43, who had one type of hrHPV infection (non HPV16/18) and the cervical cytology examination result of HSIL. Based on the above information, the model predicted an elevated risk of CIN2+. The second patient was a woman without any HPV infection. Therefore, our model predicted a very low risk of CIN2+. The third woman only had HPV16 infection and was asked by her doctor to directly undergo colposcopy without cervical cytology examination. The model predicted an elevated risk of CIN3+ for this patient. The fourth woman also had no HPV infection. Based on this, our model predicted a very low risk of CIN3+. Furthermore, the final models can be easily implemented in real-world clinical practice by

	No CIN2+ (N = 252,583)	CIN2+ (N = 1851)	P-value	Total (N = 254,434)
Age			0.06	
Median (Q1, Q3)	48.00 (42.00, 54.00)	48.00 (43.00, 54.00)		48.00 (42.00, 54.00)
Ethnicity, n (%)			0.01	
Missing	836 (0.33)	8 (0.43)		844 (0.33)
Other	2551 (1.01)	31 (1.67)		2582 (1.01)
Han	249,196 (98.66)	1812 (97.89)		251,008 (98.65)
Previous cervical cancer screening, n (%)			<0.001	
Missing	52 (0.02)	0 (0.00)		52 (0.02)
No	191,485 (75.81)	1475 (79.69)		192,960 (75.84)
Yes	61,046 (24.17)	376 (20.31)		61,422 (24.14)
Previous cervical cancer screening intervals, n (%)			<0.001	
Missing	356 (0.14)	2 (0.11)		358 (0.14)
More than three years	28,934 (11.46)	199 (10.75)		29,133 (11.45)
Less than three years	31,808 (12.59)	175 (9.45)		31,983 (12.57)
No screening	191,485 (75.81)	1475 (79.69)		192,960 (75.84)
Gynecological examination, n (%)			<0.001	
Missing	25 (0.01)	0 (0.00)		25 (0.01)
Normal	191,666 (75.88)	1201 (64.88)		192,867 (75.80)
Abnormal	60,892 (24.11)	650 (35.12)		61,542 (24.19)
Condyloma acuminata, n (%)			0.86	
Missing	25 (0.01)	0 (0.00)		25 (0.01)
No	252,543 (99.98)	1851 (100.00)		254,394 (99.98)
Yes	15 (0.01)	0 (0.00)		15 (0.01)
Trichomonas vaginalis, n (%)			<0.001	
Missing	25 (0.01)	0 (0.00)		25 (0.01)
No	251,394 (99.53)	1831 (98.92)		253,225 (99.52)
Yes	1164 (0.46)	20 (1.08)		1184 (0.47)
Vulvovaginal candidiasis, n (%)			0.81	
Missing	25 (0.01)	0 (0.00)		25 (0.01)
No	248,093 (98.22)	1821 (98.38)		249,914 (98.22)
Yes	4465 (1.77)	30 (1.62)		4495 (1.77)
Bacterial vaginosis, n (%)			0.01	
Missing	25 (0.01)	0 (0.00)		25 (0.01)
No	246,907 (97.75)	1790 (96.70)		248,697 (97.75)
Yes	5651 (2.24)	61 (3.30)		5712 (2.24)
Mucopurulent Cervicitis, n (%)			<0.001	
Missing	25 (0.01)	0 (0.00)		25 (0.01)
No	247,929 (98.16)	1770 (95.62)		249,699 (98.14)
Yes	4629 (1.83)	81 (4.38)		4710 (1.85)
Cervical polyp, n (%)			<0.001	
Missing	4688 (1.86)	58 (3.13)		4746 (1.87)
No	242,978 (96.20)	1752 (94.65)		244,730 (96.19)
Yes	4917 (1.95)	41 (2.22)		4958 (1.95)
Uterine fibroid, n (%)			0.40	
Missing	25 (0.01)	0 (0.00)		25 (0.01)
No	250,155 (99.04)	1828 (98.76)		251,983 (99.04)
Yes	2403 (0.95)	23 (1.24)		2426 (0.95)
Any of 14 hrHPV genotypes[a], n (%)			<0.001	
Negative	227,321 (90.00)	0 (0.00)		227,321 (89.34)
Positive	25,262 (10.00)	1851 (100.00)		27,113 (10.66)
Any of lrHPV genotypes[b], n (%)			<0.001	
Negative	249,876 (98.93)	1798 (97.14)		251,674 (98.92)
Positive	2707 (1.07)	53 (2.86)		2760 (1.08)

(Table 1 continues on next page)

	No CIN2+ (N = 252,583)	CIN2+ (N = 1851)	P-value	Total (N = 254,434)
(Continued from previous page)				
Any of possible hrHPV genotypes[c], n (%)			<0.001	
Negative	250,938 (99.35)	1815 (98.06)		252,753 (99.34)
Positive	1645 (0.65)	36 (1.94)		1681 (0.66)
Any of other 12 hrHPV genotypes[d], n (%)			<0.001	
Negative	229,964 (91.04)	659 (35.60)		230,623 (90.64)
Positive	22,619 (8.96)	1192 (64.40)		23,811 (9.36)
HPV16, n (%)			<0.001	
Negative	250,197 (99.06)	1042 (56.29)		251,239 (98.74)
Positive	2386 (0.94)	809 (43.71)		3195 (1.26)
HPV18, n (%)			<0.001	
Negative	251,095 (99.41)	1675 (90.49)		252,770 (99.35)
Positive	1488 (0.59)	176 (9.51)		1664 (0.65)
HPV31, n (%)			<0.001	
Negative	251,535 (99.59)	1723 (93.08)		253,258 (99.54)
Positive	1048 (0.41)	128 (6.92)		1176 (0.46)
HPV33, n (%)			<0.001	
Negative	251,118 (99.42)	1670 (90.22)		252,788 (99.35)
Positive	1465 (0.58)	181 (9.78)		1646 (0.65)
HPV35, n (%)			<0.001	
Negative	251,801 (99.69)	1793 (96.87)		253,594 (99.67)
Positive	782 (0.31)	58 (3.13)		840 (0.33)
HPV39, n (%)			<0.001	
Negative	250,409 (99.14)	1793 (96.87)		252,202 (99.12)
Positive	2174 (0.86)	58 (3.13)		2232 (0.88)
HPV45, n (%)			<0.001	
Negative	252,013 (99.77)	1820 (98.33)		253,833 (99.76)
Positive	570 (0.23)	31 (1.67)		601 (0.24)
HPV51, n (%)			<0.001	
Negative	250,000 (98.98)	1756 (94.87)		251,756 (98.95)
Positive	2583 (1.02)	95 (5.13)		2678 (1.05)
HPV52, n (%)			<0.001	
Negative	244,005 (96.60)	1433 (77.42)		245,438 (96.46)
Positive	8578 (3.40)	418 (22.58)		8996 (3.54)
HPV56, n (%)			<0.001	
Negative	251,061 (99.40)	1799 (97.19)		252,860 (99.38)
Positive	1522 (0.60)	52 (2.81)		1574 (0.62)
HPV58, n (%)			<0.001	
Negative	248,755 (98.48)	1467 (79.25)		250,222 (98.34)
Positive	3828 (1.52)	384 (20.75)		4212 (1.66)
HPV59, n (%)			<0.001	
Negative	251,353 (99.51)	1805 (97.51)		253,158 (99.50)
Positive	1230 (0.49)	46 (2.49)		1276 (0.50)
HPV66, n (%)			<0.001	
Negative	252,041 (99.79)	1835 (99.14)		253,876 (99.78)
Positive	542 (0.21)	16 (0.86)		558 (0.22)
HPV68, n (%)			<0.001	
Negative	250,350 (99.12)	1787 (96.54)		252,137 (99.10)
Positive	2233 (0.88)	64 (3.46)		2297 (0.90)
Number of genotypes with HPV infection, n (%)			<0.001	
Median (Q1, Q3)	0.00 (0.00, 0.00)	1.00 (1.00, 2.00)		0.00 (0.00, 0.00)
Cervical cytology, n (%)			<0.001	
No required[e]	224,404 (88.84)	0 (0.00)		224,404 (88.20)
No detection	431 (0.17)	105 (5.67)		536 (0.21)

(Table 1 continues on next page)

	No CIN2+ (N = 252,583)	CIN2+ (N = 1851)	P-value	Total (N = 254,434)
(Continued from previous page)				
NILM	21,210 (8.40)	267 (14.42)		21,477 (8.44)
ASC-US	3757 (1.49)	385 (20.80)		4142 (1.63)
LSIL	1723 (0.68)	232 (12.53)		1955 (0.77)
AGC-NOS	114 (0.05)	39 (2.11)		153 (0.06)
ASC-H	571 (0.23)	215 (11.62)		786 (0.31)
AGC-FN	12 (0.00)	17 (0.92)		29 (0.01)
HSIL	359 (0.14)	577 (31.17)		936 (0.37)
Cancer	2 (0.00)	14 (0.76)		16 (0.01)

Divide the participants in the development cohort into two groups according to whether CIN2+ lesions occur or not, and show the positive rates of predictors in the two groups. CIN2+, Cervical intraepithelial neoplasia grade 2 or worse; HPV, Human papillomavirus; hrHPV, High risk human papillomavirus; lrHPV, Low risk human papillomavirus; NILM, Negative for intraepithelial lesion or malignancy; ASC-US, Atypical squamous cells of undetermined significance; LSIL, Low-grade squamous intraepithelial lesion; AGC-NOS, Atypical glandular cells, not otherwise specified; ASC-H, Atypical squamous cells cannot exclude high-grade squamous intraepithelial lesion; AGC-FN, Atypical glandular cells, favor neoplastic; HSIL, High-grade squamous intraepithelial lesion. [a]The14 hrHPV genotypes refer to HPV16, HPV18, HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, HPV59, HPV66, HPV68. [b]The lrHPV genotypes refer to HPV6, HPV11, HPV42, HPV43, HPV44, HPV81. [c]The possible hrHPV genotypes refer to HPV26, HPV53, HPV70, HPV73, HPV82. [d]The other 12 hrHPV genotypes refer to HPV31, HPV33, HPV35, HPV39, HPV45, HPV51, HPV52, HPV56, HPV58, HPV59, HPV66, HPV68. [e]A negative result for high-risk HPV testing, according to the ASCCP guidelines on HR-HPV primary screening strategy, does not necessitate further cervical cytology testing.

Table 1: **Characteristics of the development cohort with CIN2+ as the endpoint.**

visiting the online website (http://cerviscreen.fjsfy.com/predict-tool). The online prediction tool is called SMART-HPV (Appendix S3).

Discussion

This study developed a machine learning prediction model for cervical cancer based on a 599,043 hrHPV full genotyping data. In five external validations across different screening settings, our model demonstrates exceptional performance in distinguishing CIN2+/CIN3+, surpassing other currently reported models.[17] Our model has good calibration and interpretability, enhancing the effectiveness of clinical practice through risk-based stratified management and interpretability. Its online tool, SMART-HPV, promotes its application in various scenarios.

We established cervical cancer prediction model employed hrHPV full genotyping and readily available test results to achieve risk prediction and stratified management for colposcopy referrals, which has great potential for its application. The 2019 ASCCP model only conducts individual detection on some hrHPV genotypes and does not perform full genotyping on all 14 high-risk types.[9] However, in addition to HPV16 and HPV18, approximately 30% of cervical cancer cases are associated with the other 12 high-risk genotypes.[28,29] Previous evidences indicates that using hrHPV full genotyping tests can provide personalized risk stratification for patients in cervical cancer screenings and improve screening efficiency.[10–13] With technological advancements, hrHPV full genotyping has been widely applied in many settings.[11,13] The National Medical Products Administration (NMPA) in China has approved at least 39 hrHPV full genotyping kits (https://www.nmpa.gov.cn/datasearch/home-index.html#category=ylqx). Unfortunately, there seems to be a lack of research on risk prediction models for cervical cancer screenings based on hrHPV full genotyping. Our model was trained and validated using cervical cancer screening data based on genotyping 14 types of hrHPV. In external validations across multiple clinical settings, it has better predictive performance for CIN2+/CIN3+ lesions than previous cervical cancer prediction models.[17,30–33] Previous studies have utilized machine learning to estimate cervical cancer risk,[30–33] their applicability in real-world cervical cancer screening has been limited due to small sample sizes or the lack of actual cervical cancer screening data, hindering the practical implementation of these models. The reported studies on cervical cancer risk prediction models lacked sufficient accuracy,[17,32] with a maximum AUROC only reaching 0.87, which is lower than our prediction model. Previous models failed to explain how to address the issue of class imbalance in cancer screening data, and the calibration of the models also needs to be improved. During our model training, we fully considered the issue of class imbalance, and through temporal validation and multiple external validations, the final XGBoost prediction model achieved an AUROC as high as 0.989 (95% CI 0.987–0.992) in external validations.

The value of our research is that have established a machine learning cervical cancer risk prediction model for settings with low screening rates. The 2019 ASCCP model introduced risk-based management in cervical cancer screening for the first time, which promoted accurate patient classification and colposcopy referrals.[9] The development cohort of the 2019 ASCCP model uses Kaiser Permanente Northern California (KPNC) data, the subjects in KPNC are demographically similar to the population in the Bay Area Metropolitan Statistical Area,

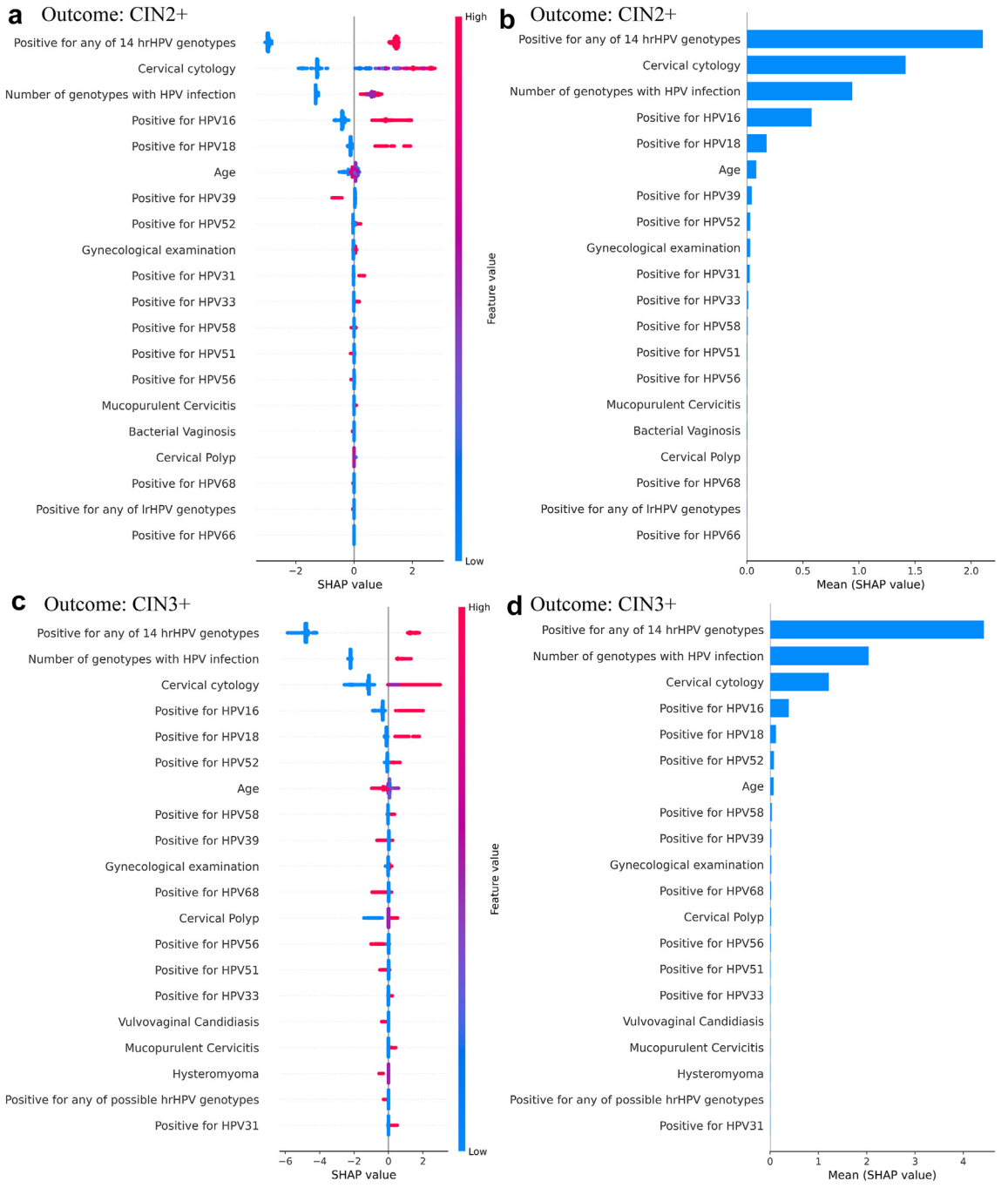

Fig. 2: **Feature importance**. To explain our model's predictors, we assessed the importance of each predictor, using Shapley values to determine their relative contributions, thereby revealing key predictors. Each point in the figure represents the Shapley (importance) value for an individual patient. The color of each point indicates the magnitude and direction of that patient's feature value. Patients with Shapley values less than zero indicate a decreased probability of being diagnosed with CIN2+/CIN3+. Furthermore, higher probabilities suggest a greater likelihood of CIN2+/CIN3+. a and b in the figure represent the results for the CIN2+ model; c and d represent the results for the CIN3+ model. CIN2+, Cervical intraepithelial neoplasia grade 2 or worse; CIN3+, Cervical intraepithelial neoplasia grade 3 or worse.

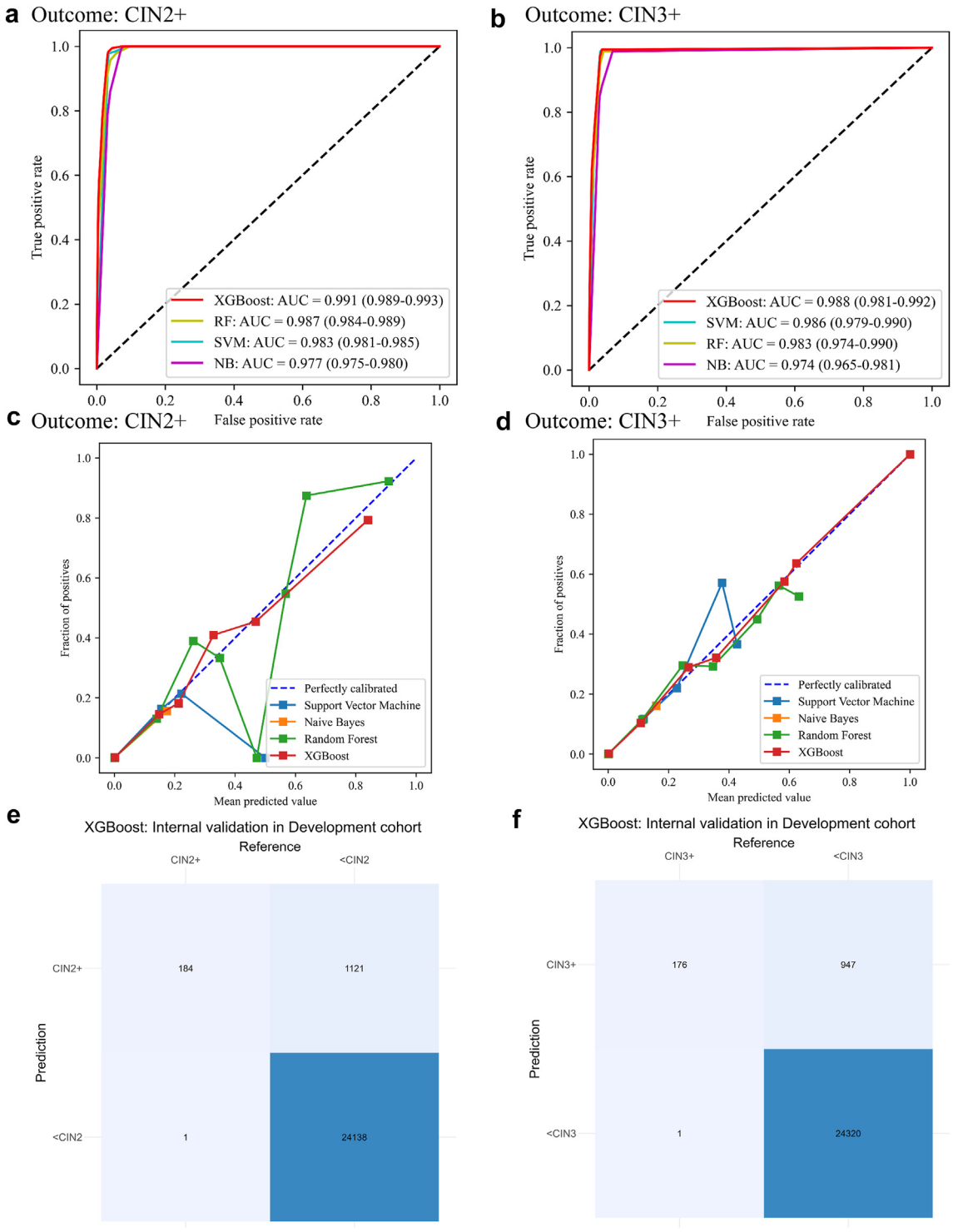

***Fig. 3*: The overall predictive performance of the four models**. a and b represent the receiver operating characteristic curve for four prediction models with CIN2+ or CIN3+ as the disease endpoints in the internal validation set, where lines of different colors represent the four trained prediction models. The discrimination of the four models in the temporal validation cohort and the external validation cohort was shown in Supplementary Table S4 and Supplementary Table S5. c and d represent the calibration curves for the four prediction models targeting the disease endpoints of CIN2+ or CIN3+ in the internal validation set, with lines of different colors indicating the four trained models. The calibration of these four predictive models in the temporal validation cohort and the external validation cohort is shown in Supplementary Table S6. e and f represent the confusion matrixes for the XGBoost models in the internal validation set, using cervical histopathology results as the gold standard. AUROC, Area under the receiver operating characteristic curve; CI, Confidence interval.

Fig. 4: **Decision curves and explainability of the models**. a and b represent the decision curves for the four predictive models with disease endpoints of CIN2+ or CIN3+. c–f show the SHAP plots of CIN2+ and CIN3+ outcomes of four patients (with and without CIN2+/CIN3+). The first patient was a woman aged 43, who had one type of hrHPV infection (non HPV16/18) and the cervical cytology examination result of HSIL. Based on the above information, the model predicted an elevated risk of CIN2+. The second and fourth patients were both women without any HPV infection. Therefore, our model predicted a very low risk of CIN2+ and CIN3+. The third woman only had HPV16 infection and was asked by her doctor to directly undergo colposcopy without cervical cytology examination.

and are considered a well-screened population with a risk of cervical cancer lower than the national average.[34] External validation cohort of the 2019 ASCCP model uses screening data from three different settings in the United States: (1) the Centers for Disease Control and Prevention's (CDC) National Breast and Cervical Cancer Early Detection Program (NBCCEDP), (2) the Onclarity HPV Trial, and (3) the Addressing the Need for Advanced HPV Diagnostic (ATHENA) study.[9] It is undeniable that the ASCCP model has been well-validated in different settings in the United States, which is also the reason why the ASCCP model is widely used in high-income settings such as the United States. Among the three external validation cohorts of the ASCCP model, the lowest cervical cancer screening rate is 69%, far higher than the global screening coverage rate of

33% among women aged 20–70 years.[7] The screening coverage rate for Chinese women aged 35–49 years is also only 33%.[7] The serious problem caused by the low screening rate is that many women who receive screening have not had cervical cancer screening before and cannot obtain previous screening results. The 2019 ASCCP guidelines use the prevalence-incidence mixture prediction model.[9] This prediction model uses logistic regression to predict the current prevalence of CIN3+ and uses proportional hazards to predict the future incidence of CIN3+ by using previous and current screening results.[9] Obtaining qualified previous and current screening results is the key to the accurate prediction of the 2019 ASCCP model.[9] But this is in contradiction with the lack of previous screening results in the low screening rate settings. These prerequisites may pose challenges to the application of the 2019 ASCCP model in global clinical settings with low screening rates. Our model was trained and validated using data of cervical cancer screening from settings with low screening rates. It has been proven through validations across various settings to effectively predict the risk of cervical cancer.

Our model was trained with great care in the selection of features, focusing on the importance of features and their availability in various clinical practices. When selecting the study population, we included 2582 individuals from minoritized ethnic groups. This reflects the representativeness of the participants in our study and also provides evidence for the use of the predictive model among minoritized ethnic. Through discussions at expert panels covering gynecology, pathology, and epidemiology, and through literature reviews, combined with the feature importance SHAP analysis, we finally included nine proven important predictors, such as hrHPV full genotyping, cytology, and gynecological examinations as model features. Previous studies primarily focused on identifying significant factors in the occurrence of cervical cancer through models such as logistic regression.[30–33] Wu Z emphasized the importance of cytological and HPV testing in identifying cervical lesions.[30] Asadi F and Kaushik M highlighted the significance of contraceptive methods, age, smoking, and other demographic characteristics in predicting cervical cancer.[31,32] Another study discovered that including HPV16/18 genotyping could enhance the accuracy of cervical cancer predictions.[17] Research has indicated that full genotyping testing of hrHPV can more accurately pinpoint the risk for CIN2+/CIN3+,[33] minimizing unnecessary colposcopy referrals and enabling more precise triage following screening. The evidence mentioned above is also why we ultimately incorporated hrHPV full genotyping into the key features of our prediction model. The outstanding performance of the model in external validations confirms that our decision was sound.

When comparing the nine predictors in our model with those in the 2019 ASCCP model, we found that the same predictors are current hrHPV and cytology test results. The difference lies in our model's inclusion of three additional predictors: hrHPV full genotyping, age, and whether the gynecological examination is abnormal. Previous studies have shown that full genotyping of 14 types of hrHPV can enhance cervical cancer screening efficiency,[10–13] and evidence has also indicated significant differences in the carcinogenic risks among these 14 types of hrHPV.[35] SHAP analysis of predictor importance further reveals substantial differences in the importance of different genotypes, leading us to include HPV16, HPV18, HPV52, and HPV39 genotyping as model predictors. The 2019 ASCCP model includes additional predictors: previous hrHPV test results and a history of histological high-grade squamous intraepithelial lesion (HSIL).[9] These predictors were initially considered in our model's candidate predictors. However, during the importance ranking of predictors, it was found that the importance of previous hrHPV test results and a history of HSIL was very low. Additionally, considering the difficulty of obtaining these predictors in low screening settings, our model ultimately excluded them. Despite excluding previous hrHPV test results and a history of HSIL as predictors, our model demonstrated good predictive performance in five external validations.

Based on the discrimination and calibration results, the overall performances of our final XGBoost models were compelling, as reflected in the decision curve analysis which illustrates clinical utility across the wide probability threshold ranges.[36] In prior studies,[17,30–33] these models have seldom been formally evaluated for clinical utility, limiting their use in practice. However, the results provided by clinical prediction models should be based on a joint decision made after considering the acceptability to clinicians and patients, which is a necessary prerequisite for translation into clinical applications.[24] By comparing with two reference strategies (all for colposcopy or none for colposcopy) and another three algorithms, DCA indicates that with thresholds lower than 0.61, the final XGBoost models demonstrated optimal net benefit in identifying cervical cancer screening patients who may progress to CIN2+/CIN3+. Additionally, the predictive results must be interpretable in accordance with real-world clinical practices. We used the SHAP algorithm to explain model prediction. Furthermore, to maximize the real-world clinical implementation of final models, we developed an online application named SMART-HPV. The accessibility and convenience of this model determine the feasibility of its clinical application; the 9 features required by the SMART-HPV model all originate from mandatory collections or tests in national cervical cancer screening of China, making them easily obtainable and interpretable in clinical practice.

The strengths of this study include separate temporal validation and external validation cohorts from six distinct settings in China. The detailed settings of the validation cohorts are aimed at comprehensively evaluating the effectiveness of our model in different environments. Specifically, in terms of residents' income levels, there are cohorts from high-income, middle-income and low-income areas; geographically, there are cohorts from the southeastern, southwestern, northwestern and central regions; considering the composition of the screening population, it includes community screened populations and opportunistic screened populations. Our models were well-calibrated and only requires nine variables that are easily obtainable in the cervical cancer screening practice. From internal to external validation, the slight decrease in discrimination suggests that overfitting is negligible, bolstering our confidence in the robustness of the models. Regarding the selection and review of features included, the clinical experts chosen for the study have extensive experience in cervical cancer diagnosis, covering disciplines such as gynecology, pathology, and laboratory medicine. The consensus reached by clinical experts, statisticians and methodologist in prediction modelling, coupled with previous literatures relating to cervical cancer screening, provides us with adequate confidence to believe that the features selection and review process was unbiased and robust.

This study has certain limitations. First, the use of smoking and oral contraceptives has been reported to potentially affect cervical cancer incidence,[37] yet these features were not collected in our study. That's because the Chinese National Cervical Cancer Screening Program does not require these features to be collected. Fortunately, our final models have stable and high performances in predicting CIN2+/CIN3+ outcomes and shows good calibration with significant net benefits. Secondly, the development cohort used for training the model includes women aged 35–64, as the Chinese National Cervical Cancer Screening targets this age group. Considering 2019 ASCCP guidelines suggest HPV genotyping screening is applicable to women aged 25–64,[9] we expanded our temporal and external validations to include women within this age range to evaluate model performance in a broader population. Third, although our models have been validated externally in six regions in China, serving as a tool for clinicians to assess whether cervical cancer screening patients should be referred for colposcopic examination still needs further research such as clinical trials and more validations from various settings worldwide. Finally, our predictive model is appropriate for cervical cancer screenings based on initial hrHPV genotyping tests because the training data from an hrHPV full genotyping initial screening. Initial screening using hrHPV genotyping have been recommended as the preferred approach in cervical cancer screening guidelines and are prevalently used in settings with low screening rates.[9,21,22] Although our model can be widely used in settings with low screening rates, it is still important to be mindful of its scope of application.

In conclusion, we developed simplified, validated high-performance prediction models for CIN2+/CIN3+. Extensive internal, temporal and external validations demonstrated that the models can be applied for guiding colposcopic referrals in cervical cancer screening across different clinical settings. The models could offer net benefits over a broad range of probability thresholds. An online application is developed, making it highly feasible for use in settings with low cervical cancer screening rates.

Contributors

P Sun, H Zou, B Dong, and Z Lu conceptualized and designed the study. T Yang, X Tuo, J Wang, S Lin, H Cai, H Cheng, X Cao, Z Zheng and H Xue contributed to data collection and specimen collection. X Huang and C Miao contributed to data verification and review. P Sun, B Dong and Z Lu verified the data and had access to the raw data. S Xu and X Liu contributed to cervical cytology and histopathology examinations. Y Zhang and Y Wang contributed to data cleaning. Z Lu, J Wang and B Dong contributed to data analysis, model training, and model validation. B Dong, Z Lu, and Y Zhang wrote the initial draft of the manuscript. All authors rigorously reviewed the manuscript. All authors have seen and approved the final report. P Sun and H Zou had final responsibility for submitting the manuscript for publication.

Data sharing statement

The individual participant data that underlie the results reported in this article will be shared after de-identification. The data will be made available starting 6 months after the publication of the article and will remain available until 3 years after its publication. These data will be shared with researchers who provide a methodologically sound application proposal. Application proposals should be sent to fmsun1975@fjmu.edu.cn and zouhuachun@fudan.edu.cn. To access the data, data requesters will need to sign a data access agreement. Requests for clarification of specific issues related to the current publication will be considered by the steering committee as long as the provision of such data does not interfere with the subsequent publications of the research team. The code that was used in this study is available from Github (https://github.com/Leslie-Lu/CerviScreen). The final models are publicly available on the website (http://cerviscreen.fjsfy.com/predict-tool/).

Declaration of interests

All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (grant no. 82271658), the Major Scientific Research Program for Young and Middle-aged Health Professionals of Fujian Province, China (grant no. 2021ZQNZD011), the Fujian Province Central Government-Guided Local Science and Technology Development Project (grant no. 2023L3019), the Fujian Province's Third Batch of Flexible Introduction of High-Level Medical Talent Teams (grant no. TD202307), the Fujian Provincial Science and Technology Innovation Joint Fund (grant no. 2021Y9171), the Fujian Provincial Health Technology Project (2024GGA061) and Fujian Provincial Natural Science Foundation of China (grant no. 2021J01403 and 2022J011040). The views expressed in the manuscript represent only those of the author and not necessarily those of the National Natural Science Foundation of China, the Fujian Provincial Department of Science and Technology, and the Fujian Provincial Health Commission. We wish to thank everyone involved in making this study a success, including all the cervical cancer screening patients and all the staff who participated in the screening.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2025.101480.

References

1 Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229–263.
2 Shahmoradi Z, Damgacioglu H, Clarke MA, et al. Cervical cancer incidence among US women, 2001-2019. *JAMA*. 2022;328:2267–2269.
3 Bowden SJ, Doulgeraki T, Bouras E, et al. Risk factors for human papillomavirus infection, cervical intraepithelial neoplasia and cervical cancer: an umbrella review and follow-up Mendelian randomisation studies. *BMC Med*. 2023;21:274.
4 Bouvard V, Wentzensen N, Mackie A, et al. The IARC perspective on cervical cancer screening. *N Engl J Med*. 2021;385:1908–1918.
5 Canfell K, Kim JJ, Brisson M, et al. Mortality impact of achieving WHO cervical cancer elimination targets: a comparative modelling analysis in 78 low-income and lower-middle-income countries. *Lancet*. 2020;395:591–603.
6 WHO. *Global strategy to accelerate the elimination of cervical cancer as a public health problem*. Geneva: World Health Organization; 2020. https://www.who.int/publications/i/item/9789240014107.
7 Bruni L, Serrano B, Roura E, et al. Cervical cancer screening programmes and age-specific coverage estimates for 202 countries and territories worldwide: a review and synthetic analysis. *Lancet Glob Health*. 2022;10:e1115–e1127.
8 Simms KT, Keane A, Nguyen DTN, et al. Benefits, harms and cost-effectiveness of cervical screening, triage and treatment strategies for women in the general population. *Nat Med*. 2023;29:30508.
9 Cheung LC, Egemen D, Chen X, et al. 2019 ASCCP risk-based management consensus guidelines: methods for risk estimation, recommended management, and validation. *J Low Genit Tract Dis*. 2020;24:90–101.
10 Li X, Rao X, Wei MJ, Lu WG, Xie X, Wang XY. Extended HPV genotyping for risk assessment of cervical intraepithelial neoplasia grade 2/3 or worse in a cohort study. *J Natl Compr Canc Netw*. 2022;20:906–914.e10.
11 Benevolo M, Ronco G, Mancuso P, et al. Comparison of HPV-positive triage strategies combining extended genotyping with cytology or p16/ki67 dual staining in the Italian NTCC2 study. *eBioMedicine*. 2024;104:105149.
12 Avian A, Clemente N, Mauro E, et al. Clinical validation of full HR-HPV genotyping HPV Selfy assay according to the international guidelines for HPV test requirements for cervical cancer screening on clinician-collected and self-collected samples. *J Transl Med*. 2022;20:231.
13 Dun C, Yuan M, Zhao X, Hu S, Arbyn M, Zhao F. Clinical evaluation of primary human papillomavirus (HPV) testing with extended HPV genotyping triage for cervical cancer screening: a pooled analysis of individual patient data from nine population-based cervical cancer screening studies from China. *Cancer Med*. 2024;13:e7316.
14 Zhu X, Yao Q, Dai W, et al. Cervical cancer screening aided by artificial intelligence, China. *Bull World Health Organ*. 2023;101:381–390.
15 Lotter W, Diab AR, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med*. 2021;27:244–249.
16 Hu L, Bell D, Antani S, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J Natl Cancer Inst*. 2019;111:923–932.
17 Xiao T, Wang C, Yang M, et al. Use of virus genotypes in machine learning diagnostic prediction models for cervical cancer in women with high-risk human papillomavirus infection. *JAMA Netw Open*. 2023;6:e2326890.
18 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
19 Council for International Organizations of Medical Sciences. *International ethical guidelines for health-related research involving humans*. Geneva, Switzerland: Council for International Organizations of Medical Sciences; 2016. http://www.cioms.ch. Accessed December 6, 2016.
20 Arbyn M, Simon M, de Sanjosé S, et al. Accuracy and effectiveness of HPV mRNA testing in cervical cancer screening: a systematic review and meta-analysis. *Lancet Oncol*. 2022;23:950–960.
21 Fontham ETH, Wolf AMD, Church TR, et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. *CA Cancer J Clin*. 2020;70:321–346.
22 Abu-Rustum NR, Yashar CM, Arend R, et al. NCCN guidelines® insights: cervical cancer, version 1.2024. *J Natl Compr Canc Netw*. 2023;21:1224–1233.
23 Andersen KM, Bates BA, Rashidi ES, et al. Long-term use of immunosuppressive medicines and in-hospital COVID-19 outcomes: a retrospective cohort study using data from the National COVID Cohort Collaborative. *Lancet Rheumatol*. 2022;4:e33–e41.
24 Zou H, Lu Z, Weng W, et al. Diagnosis of neurosyphilis in HIV-negative patients with syphilis: development, validation, and clinical utility of a suite of machine learning models. *eClinicalMedicine*. 2023;62:102080.
25 Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016:785–794.
26 Al'Aref SJ, Maliakal G, Singh G, et al. Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry. *Eur Heart J*. 2020;41:359–367.
27 Zhang J, Ma X, Zhang J, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J Environ Manage*. 2023;332:117357.
28 Wang J, Elfström KM, Lagheden C, et al. Impact of cervical screening by human papillomavirus genotype: population-based estimations. *PLoS Med*. 2023;20:e1004304.
29 Malagón T, Franco EL, Tejada R, Vaccarella S. Epidemiology of HPV-associated cancers past, present and future: towards prevention and elimination. *Nat Rev Clin Oncol*. 2024;21:522–538.
30 Wu Z, Li T, Han Y, et al. Development of models for cervical cancer screening: construction in a cross-sectional population and validation in two screening cohorts in China. *BMC Med*. 2021;19:197.
31 F A, C S, L A. Supervised algorithms of machine learning for the prediction of cervical cancer. *J Biomed Phys Eng*. 2020;10:513–522.
32 Kaushik M, Chandra Joshi R, Kushwah AS, et al. Cytokine gene variants and socio-demographic characteristics as predictors of cervical cancer: a machine learning approach. *Comput Biol Med*. 2021;134:104559.
33 Wong OGW, Ng IFY, Tsun OKL, Pang HH, Ip PPC, Cheung ANY. Machine learning interpretation of extended human papillomavirus genotyping by onclarity in an Asian cervical cancer screening population. *J Clin Microbiol*. 2019;57:e00997–e1019.
34 Manos MM, Kinney WK, Hurley LB, et al. Identifying women with cervical neoplasia: using human papillomavirus DNA testing for equivocal Papanicolaou results. *JAMA*. 1999;281:1605–1610.
35 Wei F, Georges D, Man I, Baussano I, Clifford GM. Causal attribution of human papillomavirus genotypes to invasive cervical cancer worldwide: a systematic analysis of the global literature. *Lancet*. 2024;404:435–444.
36 Cooray SD, Boyle JA, Soldatos G, et al. Development, validation and clinical utility of a risk prediction model for adverse pregnancy outcomes in women with gestational diabetes: the PeRSonal GDM model. *eClinicalMedicine*. 2022;52:101637.
37 Li XY, Li G, Gong TT, et al. Non-genetic factors and risk of cervical cancer: an umbrella review of systematic reviews and meta-analyses of observational studies. *Int J Public Health*. 2023;68:1605198.