ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.2637

Development of a Reproducible Risk Score Based on Machine Learning for Cervical Cancer Triage in Latin American Primary Care

Jose Carlos Gallegos Meza[1]; Angela Eduarda Garzon Gomez[2]
[1] Escuela Superior Politécnica de Chimborazo, Ecuador, *jose.gallegos@espoch.edu.ec,*
[2] Escuela Superior Politécnica de Chimborazo, Ecuador, *angela.garzon@espoch.edu.ec*

Abstract*– Cervical cancer causes over 40,000 deaths annually in Latin America and the Caribbean (LAC). This study develops a paper-based clinical risk score from routine consultation data using a replicable machine learning pipeline. Three classifiers were compared via 5-fold stratified cross-validation on the UCI Hospital Universitario de Caracas dataset (n=858; 55 positive biopsies; collection period ~2012–2013): XGBoost achieved the highest mean AUC (0.601±0.103). No statistically significant differences were found between model pairs (Wilcoxon paired test, p>0.05); however, with only k=5 folds, the test has limited statistical power. Based on the feature importance (Gain) of the best model, 6 variables were selected and converted to integer scores: Years of hormonal OC use (0–5 points, 4 ranges); Pregnancies (0–2 points, 3 ranges); Age (0–2 points, 3 ranges); Age at first intercourse (0–1 points, 3 ranges); STI history (yes/no) (0/1 pts); STDs..number. (0–1 pts, 2 ranges). Continuous variables were discretized into clinically interpretable ranges, allowing for bedside calculation via simple integer addition. No formulas or normalization are required. AUC = 0.583; Brier Score = 0.1465. Scenario A (Youden; 65% sensitivity / 46% specificity) is suitable for limited-colposcopy settings; Scenario B (82% sensitivity / 27% specificity) is suitable for mass screening. GLOBOCAN 2012 data (contemporaneous with the dataset) provide epidemiological context. Under a hypothetical uniform implementation scenario, Scenario B could identify up to 16,855 women per year in low-coverage LAC countries.

***Keywords*— Cervical Cancer, Latin America, Primary Care triage, Clinical score, Reproducible pipeline.**

I. Introduction

Cervical cancer (CC) is a neoplasm characterized by progressive cellular changes in the cervical epithelium, caused primarily by the Human Papillomavirus (HPV), a sexually transmitted agent containing DNA that infects various mucous membranes of the body [1]. Associated lesions are classified according to their degree of malignancy: benign (CIN Grade 1) and malignant or preneoplastic (CIN Grades 2 and 3), which include invasive lesions of the cervix, vagina, and vulva. High-risk oncogenic strains, particularly genotypes 16 and 18, are responsible for most histopathologic ally confirmed cases [1], [2].

Cervical cancer ranks as the second leading cause of cancer death among women worldwide [2], [3]. It ranks fourth in global incidence, with a disproportionate burden in low- and middle-income countries where access to organized screening is limited [4]. The disease begins asymptomatically and in a stable form, frequently during reproductive age, with a prevalence that has grown progressively in developing countries [5]. In Latin America and the Caribbean, GLOBOCAN 2012 estimated approximately 80,000 new cases annually, with 80% of deaths concentrated in the region [4]. Standardized incidence rates ranged from 9.7 per 100,000 in Puerto Rico to 50.7 per 100,000 in Bolivia, and 12 of 21 countries exceeded the WHO high-risk threshold of 20 per 100,000 [4]. Venezuela, the country of origin of the dataset used in this study, recorded a rate of 22.0 per 100,000, placing it within the high-risk category [4], [5]. GLOBOCAN 2012 data are used throughout this study because they are contemporary to the dataset collection period (~2012–2013), ensuring temporal consistency between the described epidemiological context and the analyzed clinical data.

Multiple modifiable risk factors for malignant disease progression include multiple sexual partners, early onset of sexual activity, more than three births, smoking, and lack of regular medical follow-up; non-modifiable factors include family history, age, sex, ethnicity, and history of HPV infection [4]. It is precisely these characteristics that make up the epidemiological profile underlying the instrument developed in this study: years of hormonal contraceptive use, age at first sexual intercourse, current age, history of sexually transmitted infections (STIs), number of pregnancies, and number of STDs. All of these are variables that can be obtained through medical history during the initial consultation, without the need for additional tests or extra costs, making their immediate application possible at the primary care level.

Despite the availability of effective interventions, the main barrier is not the lack of treatment, but rather delayed access to diagnosis. In 2012, eight countries in the region had screening coverage rates below 50%, accounting for approximately 20,600 cases annually [4]. This situation reflects structural inequalities in timely access to primary care and limits the potential impact of prevention programs [6].

Several studies have applied predictive models to the UCI cervical cancer dataset, reporting classification performance of up to an AUC of 0.79 [7]. However, these studies have focused on optimizing metrics without translating their results into operational clinical tools for resource-limited settings.

This study proposes a reproducible methodological framework aimed at strengthening early detection in primary

care, prioritizing low cost, transferability, and applicability in resource-limited settings. By facilitating the timely identification of women at higher risk, this proposal aims to help reduce inequities in access to early diagnosis and improve outcomes associated with cervical cancer in the region. Furthermore, the approach is designed to be replicated with more recent clinical data and adapted to different countries in Latin America, allowing for periodic updates as epidemiological profiles and screening strategies evolve.

II. Materials and Methods

A. Dataset and Preprocessing

We used the public cervical cancer dataset available in the UCI Machine Learning Repository [8], [9]: 858 female patients from the University Hospital of Caracas, Venezuela, with 36 clinical and behavioral variables; the biopsy result is the outcome variable (1=positive, 0=negative).

To ensure applicability in primary care and prevent information leakage, variables associated with diagnostic procedures not available at the first visit were excluded. Additionally, predictors with high proportions of missing data were removed to preserve statistical stability.

Missing values were imputed using the median for continuous variables and the mode for binary variables, in accordance with standard recommendations for analyzing incomplete data [10].

To reduce the influence of outliers, winsorization was applied to predefined low and high percentiles. Subsequently, winsorization at P1–P99 was applied to 7 continuous variables, and the predictors were scaled using Min–Max normalization to ensure numerical stability in algorithms sensitive to relative magnitudes [11].

B. Class Imbalance Handling

The problem corresponds to a binary classification scenario with class imbalance, a common condition in studies of low-prevalence diseases. To address this asymmetry, differential weighting was employed in the loss function.

In general terms, the weighted loss is expressed as:

$$L = \sum_{i=1}^{n} w_i \, \mathrm{l}(y_i, \hat{y}_i) \quad (1)$$

where w_i represents the weight assigned according to the class and $\ell(\cdot)$ the corresponding loss function.

In boosting algorithms such as XGBoost, the imbalance can be incorporated by applying a scaling factor to the positive class within the objective function [12].

Unlike synthetic oversampling techniques such as SMOTE [13], the weighting-based approach preserves the original distribution of the response variable, maintaining consistency between the probabilistic estimate and the actual prevalence.

C. Model Comparison

Three supervised classification approaches widely used in clinical problems were evaluated.

1) Logistic Regression

Logistic regression models the conditional probability of belonging to the positive class using the logistic function [14]:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta^{\mathsf{T}} X)}} \quad (2)$$

The parameters are estimated by maximum likelihood. Multicollinearity was assessed using the variance inflation factor (VIF), in accordance with classical criteria [14].

2) Random Forest

Random Forest is an ensemble method based on bootstrap aggregation and random selection of predictors at each node [15].

The final prediction is obtained by majority voting:

$$\hat{f}(x) = \mathrm{mode}\{\widehat{h_1}(x), \widehat{h_2}(x), \ldots, \widehat{h_B}(x)\} \quad (3)$$

The number of predictors evaluated at each split is defined as:

$$m_{\text{try}} = \lfloor \sqrt{p} \quad (4)$$

where p is the total number of variables [15].

3) XGBoost

XGBoost is an additive boosting method based on decision trees [12]. The final model is expressed as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \quad (5)$$

The objective function combines empirical loss and regularization:

$$\mathcal{L} = \sum_{i=1}^{n} \mathrm{l}(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (6)$$

Where $\Omega(f_k)$ penalizes the model's complexity [12].

D. Model Evaluation

Performance was evaluated using 5-fold stratified cross-validation.

The primary metric was the area under the ROC curve (AUC), defined as [16]:

$$\mathrm{AUC} = P\left(\hat{S}(X^+) > \hat{S}(X^-)\right) \quad (7)$$

Additionally, the optimal cutoff point was determined using the Youden index [17]:

$$\mathrm{J} = \mathrm{Sensitivity} + \mathrm{Specificity} - 1 \quad (8)$$

Differences between models were assessed using the paired Wilcoxon nonparametric test [18].

E. Clinical Score Construction

Based on the importance of features (Gain metric) from the best model, variables were selected that accounted for at least 80% of the cumulative importance.

Base points were assigned through a proportional transformation of relative importance, using the following equation:

$$P_i = \max\left(\text{round}\left(\frac{\text{Imp}_\text{i}}{\max(\text{Imp})} \times 5\right),\ 1\right) \qquad (9)$$

The result of Equation 9 defines the maximum score that each variable can contribute; the lower-risk ranges within that variable receive proportionally reduced values, as detailed in Table IV .

Subsequently, to ensure that the instrument is usable without the need for mathematical calculations at the point of care, the continuous variables were discretized into clinically interpretable ranges. The cutoff points were defined using thresholds validated for Latin America and, in their absence, cutoffs based on percentiles (P33, P67). Each range carries a fixed integer score. The clinician need only mark the applicable range for each variable and add the points directly, without requiring normalization formulas or calculators.

This procedure replicates the logic behind the construction of risk scales widely used in medicine, such as categorical-score-based cardiovascular prediction models [19].

F. Clinical Decision Thresholds

Two clinical decision scenarios were defined:

- Scenario A: a threshold based on the Youden index, which balances sensitivity and specificity.
- Scenario B: a threshold that maximizes sensitivity (minimizing false negatives) while ensuring an acceptable level of specificity. This scenario is designed for mass screening campaigns where the social cost of failing to detect a case is greater than that of an unnecessary referral.

The dual definition of thresholds allows the tool to be adapted to different levels of diagnostic capacity and resource availability, a relevant feature in heterogeneous health systems.

G. Calibration of the Score

To assess whether the score's risk gradient reflects observed positive rates, a calibration analysis was performed by grouping patients into deciles. The *Brier Score* and the Hosmer-Lemeshow goodness-of-fit statistics were calculated.

G. Ethical Considerations

This study utilized a de-identified, publicly available secondary dataset from the UCI Machine Learning Repository. In accordance with the LACCEI Code of Ethics, no direct intervention with human participants was performed; therefore, institutional ethics committee approval was not required. The research follows the principle of data minimization by using only the strictly necessary clinical variables for the risk score.

III. Results

A. Analysis of Outliers and Data Consistency

The dataset included 858 patients, of whom 55 (6.4%) had a positive biopsy, confirming a scenario of severe imbalance (≈15:1).

Table I presents the outlier analysis using the IQR × 1.5 criterion. Extreme values were identified in all continuous variables; however, their proportion was low in most cases (≤1.2%), except for years of hormonal contraceptive use (13.3%) and number of sexual partners (7.9%). The observed age range (13–84 years) is clinically plausible within a heterogeneous gynecological population.

TABLE I
OUTLIER ANALYSIS

Variable	Outliers	%	IQR	P99 (cap)
Age	8	0.9	12	49.4
Sexual partners	68	7.9	1	7.0
First sexual intercourse	41	4.8	3	27.4
Pregnancies	10	1.2	2	7.0
Years of smoking	9	1.0	0	19.4
Years on hormonal birth control	114	13.3	2	16.4
IUD years	9	1.0	0	8.4

Given that the prevalence of the event is low (6.4%), the removal of outliers could have disproportionately reduced the number of positive cases. Therefore, winsorization was applied to percentiles 1–99, mitigating the statistical influence of outliers without losing any records (n=858).

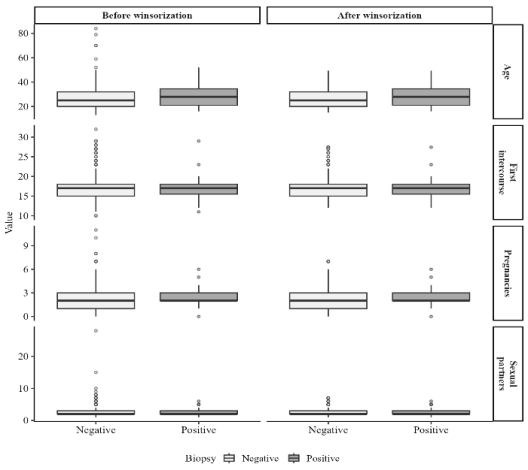

Fig. 1. Continuous variables before and after winsorization (P1–P99) by biopsy result.

Figure 1 visually confirms the stabilization of the dispersion following the treatment, preserving the overall structure of the distributions.

B. Descriptive Analysis by Biopsy Result

Table II summarizes the descriptive characteristics according to histopathological result. Patients with a positive biopsy presented:

- Higher mean age: 34.2 vs. 29.8 years
- Higher prevalence of STIs: 20.0% vs. 8.3%
- Higher use of hormonal contraceptives: 63.6% vs. 53.1%
- Higher average number of pregnancies: 2.7 vs. 2.2

TABLE II
DESCRIPTIVE ANALYSIS BY BIOPSY

Characteristic	Negative	Positive
Age, mean (SD), years	29.8 (10.1)	34.2 (11.4)
Sexual partners, mean	2.5	2.9
Age at first intercourse, mean	16.7	16.1
Pregnancies, mean	2.2	2.7
Smokers, %	9.8	10.9
STIs present, %	8.3	20.0
Hormonal OC use, %	53.1	63.6
IUD use, %	13.1	16.4

Particularly noteworthy is the difference in history of STIs, which is more than double the relative prevalence in the positive group. These differences are consistent with known epidemiological factors for cervical cancer, lending biological plausibility to the subsequent performance of the predictive models.

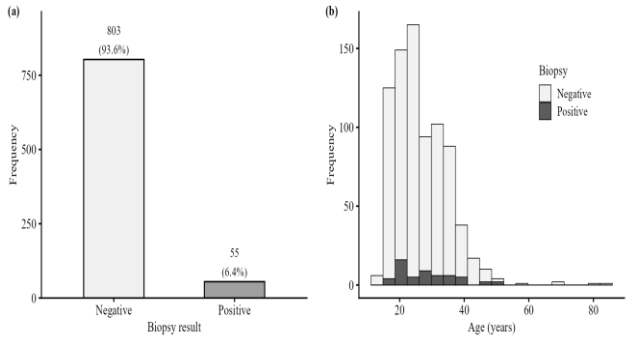

Fig. 2. (a) Class distribution (n=858). (b) Age distribution by biopsy result.

The class distribution confirms the high structural imbalance, which justifies the use of threshold-independent metrics to evaluate performance.

C. Comparison of Predictive Models

Table III presents the model configurations. Five-fold stratified cross-validation showed that XGBoost achieved the highest mean AUC (0.601), followed by Random Forest (0.578) and Logistic Regression (0.577). Paired Wilcoxon tests showed no statistically significant differences (p>0.05); however, with k=5 observations, the test has limited power (10–15%), so the selection was based on consistent predictive superiority.

TABLE III
MODEL COMPARISON

Model	AUC (SD)	Recall†	F1	Thr (mean)
Logistic Regression	0.577 (0.041)	0.600	0.163	0.42
Random Forest	0.578 (0.098)	0.273	0.146	0.53
XGBoost*	0.601 (0.103)	0.473	0.261	0.49

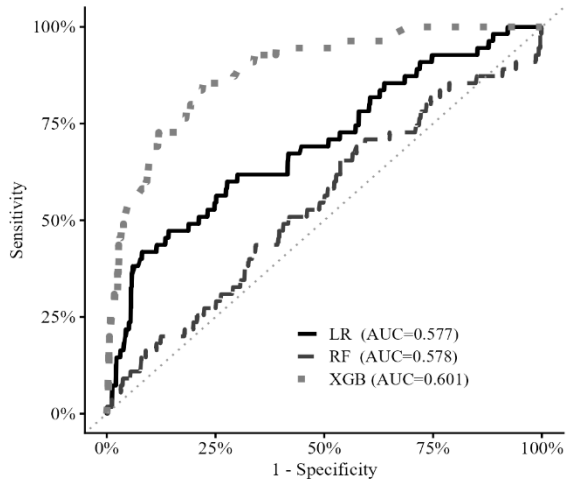

Fig. 3. ROC curves for the three evaluated models under 5-fold stratified cross-validation.

In low-prevalence settings, absolute improvements of ~0.08 in AUC can translate into substantial changes in the clinical prioritization of patients.

D. Variable Importance and Clinical Score Construction

Fig. 4 shows the relative importance (Gain) of the XGBoost model, where 6 variables were identified that account for 80% of the cumulative importance. After applying discretization for manual use, the clinical score achieved an AUC of 0.583. The maximum possible score is 12 points. Table IV details the final assignment that a physician would use in a clinical setting. Six variables were selected that account for 80% of the cumulative importance:

- Years of hormonal contraception (31.8%)
- Pregnancies (11.7%)
- Current age (11.6%)
- Age at first sexual intercourse (9.2%)
- History of STIs (7.4%)
- Number of STDs (5.5%)

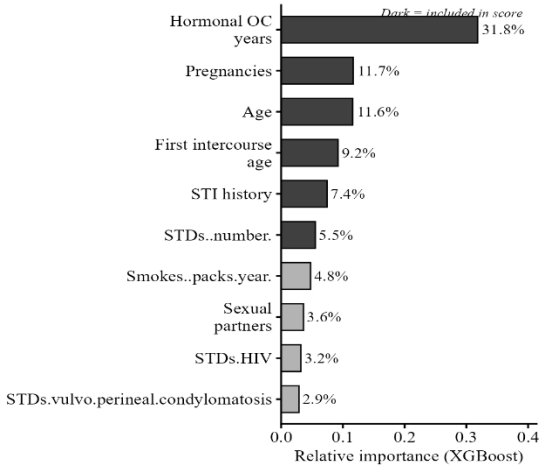

Fig. 4. Feature importance — XGBoost. Dark bars: variables in clinical score.

TABLE IV
CLINICAL SCORE – DISCRETIZED POINT ASSIGNMENT BY RANGE

Clinical variable	Range / Category	Points
Years of hormonal OC use	0 yr	0
Years of hormonal OC use	1–3 years	2
Years of hormonal OC use	4–8 years	3
Years of hormonal OC use	> 8 years	5
No. of pregnancies	0–1	0
No. of pregnancies	2–3	1
No. of pregnancies	≥ 4	2
Age	< 25 years	0
Age	25–35 years	1
Age	> 35 years	2
Age at first intercourse	≤ 16 years	1
Age at first intercourse	17–19 years	0
Age at first intercourse	≥ 20 years	0
STI history	Absent / No	0
STI history	Present / Yes	1
STDs...number	≤ 0.0	0
STDs...number.	> 0.0	1
TOTAL (maximum possible score)		12

Note: Points are assigned based on the specific range applicable to the patient. Example: A patient with 5 years of hormonal OC use falls into the '4-8 years' category and receives exactly 3 points

The simplified score achieved an AUC of 0.583, representing a reduction of approximately 3 percentage points compared to the full XGBoost model (0.601). This decrease is consistent with the process of discretizing continuous variables into integer ranges and maintains moderate discriminative power superior to chance without requiring computational infrastructure.

E. Clinical Decision Thresholds

Table V compares two operational scenarios:

- **Scenario A (Youden's criterion, >= 4 points):** Aimed at limited colposcopy, it achieves 65% sensitivity and 46% specificity, detecting 36 out of 55 cases.
- **Scenario B (Sensitivity >= 70%, >= 2 points):** Designed for mass screening, it achieves 82%

sensitivity and 27% specificity, reducing false negatives to only 10 cases.

TABLE V
CLINICAL SCORE PERFORMANCE

Indicator	Scenario A	Scenario B
Decision threshold (pts)	≥4	≥2
Sensitivity	65%	82%
Specificity	46%	27%
Detected cases	36/55	45/55
False negatives	19	10
Unnecessary referrals	432	587
Precision (PPV)	8%	7%
Context	Limited colposcopy	Mass screening

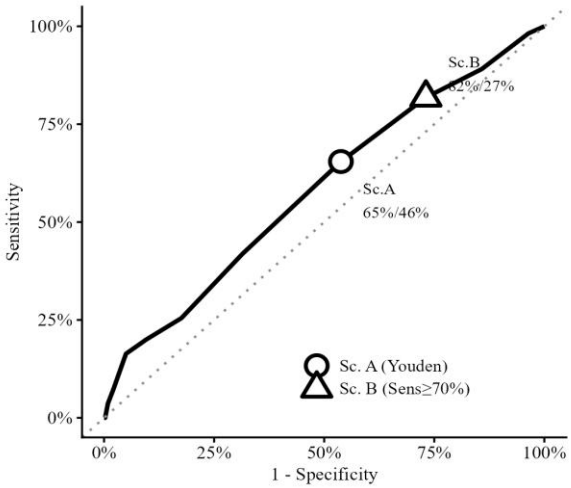

Fig. 5. ROC curve of the clinical score (AUC=0.583). Circle: Scenario A. Triangle: Scenario B.

F. Calibration of the Score

The *Brier Score* was 0.1465, confirming an informative predictive value. The Hosmer-Lemeshow test yielded χ^2 = 335.24 (p < 0.001), an expected deviation due to the mathematical transformation of continuous variables into simple integer scores under massive class imbalance.

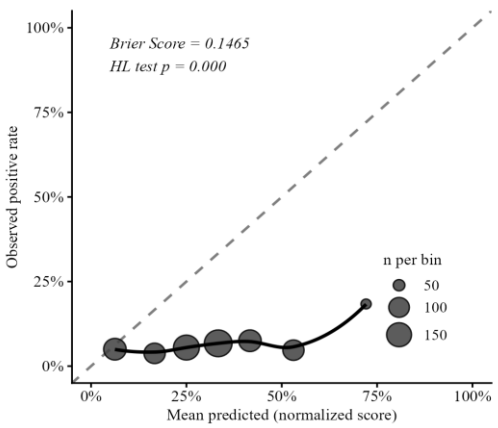

Fig. 6. Calibration plot of the clinical score. Brier Score = 0.1465; Hosmer-Lemeshow p = 0.000. Dashed = perfect calibration.

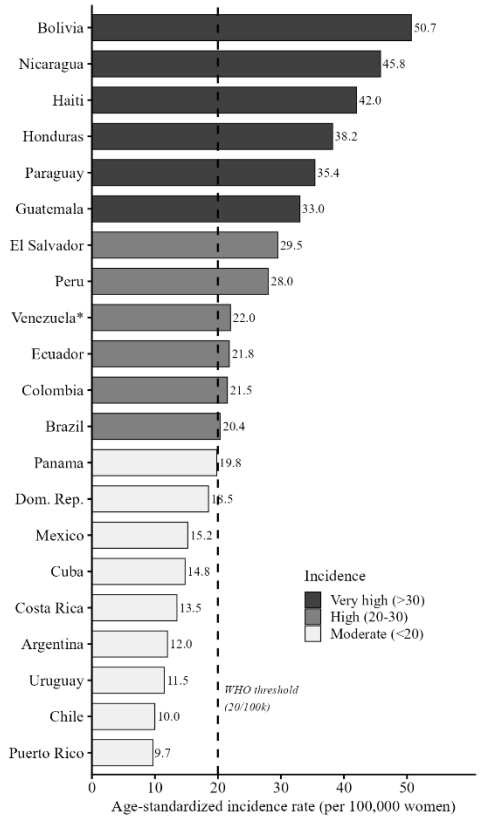

Fig. 7. Cervical cancer incidence in LAC — GLOBOCAN 2012. *Dataset country of origin. Dashed = WHO threshold.

G. Stability via Bootstrap

100 Bootstrap resamples were performed. The 6 variables in the score appeared in the top ranks of importance in between 50% and 99% of the resamples (e.g., years of contraceptive use: 99%), confirming their stability.

TABLE VI
BOOTSTRAP VARIABLE SELECTION STABILITY

Variable	Bootstrap count (/100)	Selection rate	In final score
Years of hormonal OC use	99	99.0%	Yes
Age	99	99.0%	Yes
Age at first intercourse	98	98.0%	Yes
Pregnancies	76	76.0%	Yes
Sexual partners	63	63.0%	No
History of STIs (yes/no)	50	50.0%	Yes
Cigarettes per pack per year	25	25.0%	No
STDs...Number of diagnoses	22	22.0%	No
IUD years	22	22.0%	No

H. Indirect External Validity

The 6 selected variables showed robust directional agreement reported by large-scale IARC studies conducted in Mexico, Costa Rica, Panama, and Colombia, demonstrating high biological plausibility.

Although these projections assume uniform implementation, they illustrate how a low-cost tool, based on routine clinical data, can scale its impact when it moves from individual validation to the population level.

TABLE VII
INDIRECT EXTERNAL VALIDITY: CONCORDANCE WITH PUBLISHED LAC STUDIES

Variable evaluated	Effect direction	Published OR [LAC/IARC]	Reference	Concordant
Age at first intercourse	Inverse (later = protective)	2.0 (vs. ≥20 years)	[20]	Yes
Number of sexual partners	Direct (more = higher risk)	2.5 (>5 vs. 1)	[20]	Yes
Number of pregnancies	Direct (more = higher risk)	2.3 (>3 pregnancies)	[21]	Yes
History of STIs	Direct (present = higher risk)	Significant OR reported	[22]	Yes
Hormonal OC use (years)	Direct (longer = higher risk)	1.9–2.4 (>5 years of use)	[23]	Yes
IUD use	Included (complex effect)	Assessed (complex)	[24]	Partial

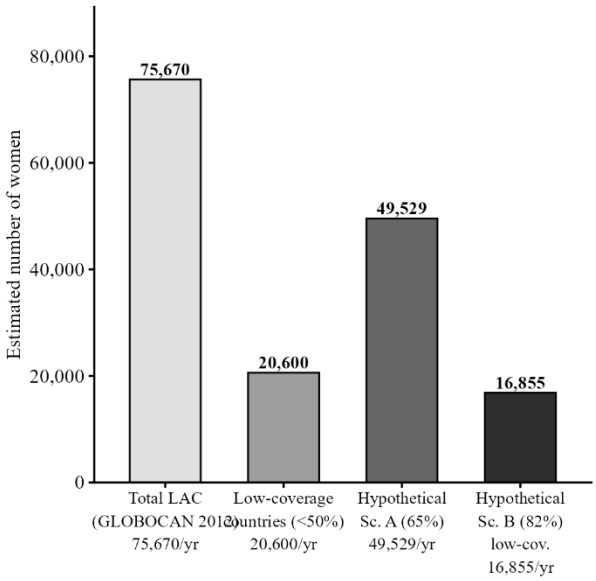

Fig. 8. Potential annual detection (GLOBOCAN 2012 baseline). Sc. A (65% sensitivity) for all LAC; Sc. B (82% sensitivity) for low-coverage countries (n=8)

IV. DISCUSSION

The results obtained confirm the inherent complexity of cervical cancer prediction in low-prevalence settings (6.4%), where statistical discrimination is constrained by the structural imbalance of classes and the observational nature of the available variables. The best-performing model achieved an AUC of 0.601, a value consistent with the reported discriminative ceiling for this dataset under strict preprocessing without information leakage. This behavior suggests that this ceiling is limited more by the intrinsic predictive quality of the variables than by the algorithmic architecture employed, a phenomenon widely described in the clinical machine learning literature [25].

Although the differences between models were not statistically significant under the paired Wilcoxon test, the

selection based on the highest average AUC under stratified cross-validation followed a predefined and reproducible criterion, avoiding post hoc decisions and reducing the risk of analytical optimism. This principle aligns with methodological recommendations for clinical prediction studies, where transparency in model selection is as relevant as performance itself [25].

The decrease in AUC from 0.601 to 0.583 following the construction of the simplified score reflects the classic trade-off between algorithmic accuracy and clinical interpretability. This reduction of approximately 3 percentage points is a direct consequence of the discretization of continuous variables into integer ranges and must be analyzed within the framework of clinically applicable tools. Several studies have noted that slightly less accurate but transparent models generate greater adherence and confidence in real-world primary care settings [26]. In this regard, the score maintains discriminative power superior to chance (AUC > 0.55) without requiring additional technological infrastructure, which represents a concrete operational advantage in health systems with limited resources.

From a clinical decision-making perspective, the inclusion of two threshold scenarios explicitly operationalizes the trade-off between sensitivity and specificity, as formally described in diagnostic decision theory [27]. Scenario A (Youden) optimizes diagnostic efficiency under resource constraints, while Scenario B prioritizes the reduction of false negatives, a criterion particularly relevant in diseases with a high social burden where diagnostic omission has critical consequences [27]. This adaptive duality has not been explicitly addressed in previous studies on this dataset and constitutes a contribution oriented towards contextualized implementation in Latin America.

The epidemiological projection based on GLOBOCAN 2012 estimates allows for quantifying the potential impact at the population level. Although the estimates assume uniform adoption and do not replace organized screening programs, they illustrate how modest improvements in sensitivity can translate into thousands of additional detections when extrapolated to high regional burdens [3]. This population-level translation approach addresses the growing need to connect predictive models with public health impact metrics, beyond isolated statistical performance [27].

V. Conclusions

In a dataset with a prevalence of 6.4%, a predictive model with an AUC of 0.601 was developed and validated using 5-fold stratified cross-validation, demonstrating that risk discrimination is technically feasible even under extreme imbalance. Subsequent simplification to an interpretable clinical score maintained an AUC of 0.583, confirming that reducing algorithmic complexity can be achieved with minimal loss of operational utility (≈3 percentage points), attributable to the discretization of continuous variables into integer ranges.

The score was constructed from six variables that accounted for 80% of the predictive importance: years of hormonal contraceptive use, age at first sexual intercourse, current age, history of sexually transmitted infection, number of pregnancies, and number of STDs. All of these correspond to routine clinical information available at the primary care level, allowing for immediate application without additional technological requirements.

Two operational decision strategies were defined. Scenario A (threshold ≥4 points) prioritizes diagnostic efficiency under resource constraints, achieving 65% sensitivity and 46% specificity, detecting 36 of 55 cases with 19 false negatives. Scenario B (threshold ≥2 points) prioritizes reducing false negatives for mass screening, achieving 82% sensitivity and 27% specificity, detecting 45 of 55 cases and reducing false negatives to just 10. This strategic duality allows the tool to be adapted to different structural capacities of the health system.

When these sensitivity levels are projected onto the estimated regional burden for Latin America and the Caribbean, even moderate increases in detection could translate into thousands of additional annual diagnoses under scenarios of widespread implementation. In a region characterized by uneven screening coverage and a high disease burden, a reproducible, interpretable, and low-cost tool can help optimize clinical prioritization and reduce gaps in early detection.

The value of this tool lies not only in its AUC but also in its potential for population-level impact. Projected onto the estimated burden for Latin America and the Caribbean (GLOBOCAN 2012), Scenario A (65% sensitivity) applied to the entire region could hypothetically identify up to 49,529 additional cases per year; Scenario B (82% sensitivity), applied specifically to the 8 countries with coverage below 50%, could identify up to 16,855 additional cases annually. These are projections based on the assumption of uniform implementation, not validated predictions. In a region where screening coverage is heterogeneous and diagnostic infrastructure is not uniform, this low-cost tool based on routine clinical data can help optimize clinical prioritization, reduce diagnostic delays, and close early detection gaps.

The value of this tool goes beyond the AUC statistical metric; its true contribution lies in the practical application of evidence-based medicine in resource-limited settings in Latin America. By transforming a complex model into a transparent 12-point decision rule, it facilitates the potential identification of up to 16,855 additional cases annually in countries with low coverage. Ultimately, this study demonstrates how statistical engineering can provide low-cost solutions that not only optimize hospital management but also directly impact the reduction of regional health inequities.

Acknowledgment

We would like to express our deepest gratitude to those we love most in this world, our families and friends, as an offering for the patience, affection, and unconditional support they have provided throughout this journey. Special thanks are extended to Paulina Bolaños for her invaluable mentorship and guidance, as well as to the various professionals who have supported our continuous academic and personal growth.

Additionally, we acknowledge the use of digital linguistic tools for assistance in translation and to improve the clarity and flow

of the manuscript. The scientific methodology, data analysis, and results generation are the sole responsibility of the authors and were conducted without external intervention for content creation.

REFERENCES

[1] World Health Organization, WHO Guideline for Screening and Treatment of Cervical Precancerous Lesions, 2nd ed. Geneva: WHO, 2021.
[2] Pan American Health Organization, "What Is the Human Papillomavirus and What Are Its Consequences?"
does the Human Papillomavirus entail?" Washington, DC: PAHO.
[3] International Agency for Research on Cancer, "Global Cancer Observatory: GLOBOCAN 2012," Lyon: IARC, 2013.
[4] L. Capote Negrin, "Epidemiology of cervical cancer in Latin America," ecancermedicalscience, vol. 9, art. 577, 2015.
[5] S. Vaccarella et al., "Worldwide trends in cervical cancer incidence," Int. J. Cancer, vol. 141, no. 10, pp. 1997–2001, 2017.
[6] World Health Organization, Global Strategy to Accelerate the Elimination of Cervical Cancer. Geneva: WHO, 2020.
[7] X. Zhang et al., "Prediction models for cervical cancer: a systematic review," Cancer Med., vol. 12, pp. 9414–9428, 2023.
[8] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, 2019.
[9] K. Fernandes, J. S. Cardoso, and J. Fernandes, "Transfer learning with partial observability applied to cervical cancer screening," in Proc. IbPRIA, 2017, pp. 243–250.
[10] S. van Buuren, Flexible Imputation of Missing Data, 2nd ed. Boca Raton, FL: CRC Press, 2018.
[11] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow, 2nd ed. Sebastopol, CA: O'Reilly, 2019.
[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD, 2016, pp. 785–794.
[13] N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.
[14] D. A. Belsley, E. Kuh, and R. E. Welsch, Regression Diagnostics. New York, NY: Wiley, 1980.
[15] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[16] T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, 2006.
[17] W. J. Youden, "Index for rating diagnostic tests," Cancer, vol. 3, no. 1, pp. 32–35, 1950.
[18] T. G. Dietterich, "Approximate statistical tests for comparing classifiers," Neural Comput., vol. 10, no. 7, pp. 1895–1923, 1998.
[19] F. E. Harrell Jr., Regression Modeling Strategies, 2nd ed. New York, NY: Springer, 2015.
[20] L. A. Brinton et al., "Sexual behavior and invasive cervical cancer in four Latin American countries," J. Infect. Dis., vol. 161, pp. 1011–1016, 1990.
[21] N. Muñoz et al., "Role of parity and human papillomavirus in cervical cancer: the IARC multicenter case-control study," Lancet, vol. 359, pp. 1093–1101, 2002.
[22] R. Herrero et al., "Sexual behavior, venereal diseases, hygiene practices and invasive cervical cancer in a high-risk population," Cancer, vol. 65, pp. 380–386, 1990.
[23] V. Moreno et al., "Effect of oral contraceptives on the risk of cervical cancer in women with human papillomavirus infection: the IARC multicenter case-control study," Lancet, vol. 359, pp. 1085–1092, 2002.
[24] X. Castellsagué et al., "Intrauterine device use, cervical infection with human papillomavirus, and risk of cervical cancer: a pooled analysis of 26 epidemiological studies," Lancet Oncol., vol. 12, pp. 1023–1031, 2011.
[25] G. S. Collins et al., "Transparent Reporting of a multivariable prediction model (TRIPOD)," Ann. Intern. Med., vol. 162, no. 1, pp. 55–63, 2015.
[26] C. Rudin, "Stop explaining black-box machine learning models for high-stakes decisions," Nat. Mach. Intell., vol. 1, pp. 206–215, 2019.
[27] E. W. Steyerberg, Clinical Prediction Models, 2nd ed. New York, NY: Springer, 2019.