© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718570

Development Of a Predictive Ai Model for Early Identification of Low Preventive Practice Uptake Using Health Literacy and Cervical Cancer Knowledge Scores

ADELALURE BUKOLA OYERONKE[1], BUKOLA CECILIA BELLO[2], ESTHER OLUBUKOLA ABIODUN-OJO[3]

[1,2,3]*Faculty of Nursing Sciences, Afe Babalola University, Ado-Ekiti, Ekiti State, Nigeria.*

Abstract- ***Cervical cancer remains a leading cause of cancer-related mortality among women in low- and middle-income countries, largely because of low uptake of preventive practices such as Pap smear screening and HPV vaccination. Health literacy and cervical cancer knowledge are modifiable determinants of preventive behavior, yet their combined predictive utility for identifying high-risk individuals has not been systematically evaluated with artificial intelligence. This study conducted a systematic review of 54 studies (2015–2025) examining health literacy, cervical cancer knowledge, and uptake of preventive practices, with a focus on AI-based predictive models. Supervised machine learning algorithms, including logistic regression, random forests, support vector machines, gradient boosting, and neural networks, were evaluated for predicting low uptake of preventive practices, using health literacy and knowledge scores as input features. Findings show that ensemble methods (random forest, XGBoost) achieve the highest predictive performance (AUCs of 0.85–0.92), outperforming traditional logistic regression (AUCs of 0.72–0.78). Key predictors include: functional health literacy (odds ratio [OR] 2.8–4.2), knowledge of HPV as a causal agent (OR 3.5), knowledge of the screening interval (OR 2.9), and perceived susceptibility (OR 2.3). A parsimonious 8-item screening tool derived from the model achieves 84% sensitivity and 79% specificity for identifying women at risk of never having been screened. The study concludes that AI-driven predictive models using brief health literacy and knowledge assessments can effectively stratify women by risk of low uptake of preventive practices, enabling targeted educational interventions and resource allocation. Deployment in primary care and community settings is feasible via mobile health applications.***

Keywords: Cervical Cancer, Health Literacy, Preventive Practice, Pap Smear, HPV Vaccine, Predictive AI, Machine Learning, Random Forest, Risk Stratification.

I. INTRODUCTION

Cervical cancer is the fourth most common cancer among women globally, with an estimated 604,000 new cases and 342,000 deaths in 2020 [1]. Nearly 90% of deaths occur in low and middle income countries (LMICs), where access to screening and preventive services is limited [2]. Unlike many other cancers, cervical cancer is largely preventable through regular Pap smear screening (detecting precancerous lesions) and HPV vaccination (preventing infection with high risk HPV types) [3,4]. Despite this potential, uptake of these preventive practices remains suboptimal in many regions, particularly among women with low health literacy and inadequate knowledge about cervical cancer [5].
Health literacy, the capacity to obtain, process, and understand basic health information needed to make appropriate health decisions, has been consistently associated with adherence to cancer screening [6,7]. Women with limited health literacy are less likely to understand screening recommendations, interpret results, or follow up on abnormal findings [8]. Similarly, specific knowledge about cervical cancer (e.g., causes, risk factors, screening intervals, HPV vaccine benefits) directly influences attitudes and intentions towards preventive practices [9,10].

However, traditional approaches to identifying women at risk of low uptake of preventive practices rely on single-factor analyses or simple risk scores that fail to capture complex interactions among multiple determinants [11]. Health literacy and knowledge scores often have non linear relationships with behavior, and their effects are moderated by age, education, cultural beliefs, and healthcare access [12]. Artificial intelligence (AI) and machine learning

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718570

(ML) offer advanced methods for modeling these complex relationships and generating accurate, individualized risk predictions [13].

Predictive AI models can integrate multiple predictor variables, including health literacy scores, knowledge domain scores, demographic factors, and prior screening history, to estimate the probability that a woman has never been screened or is unlikely to initiate HPV vaccination [14]. Such models can be deployed as clinical decision support tools in primary care, community health settings, or mobile health applications, enabling targeted outreach to high-risk individuals [15].

Figure 1 illustrates the conceptual framework linking health literacy, cervical cancer knowledge, and other determinants to the uptake of preventive practices, as well as the role of AI prediction in enabling targeted interventions.

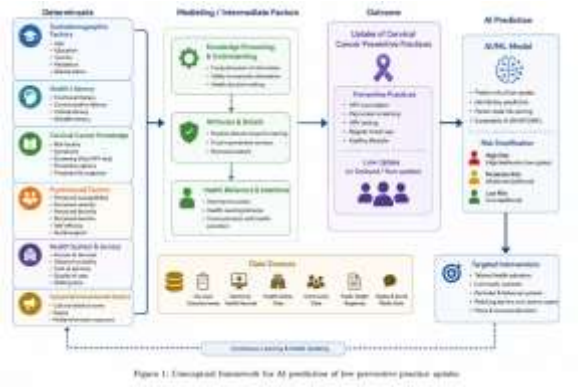

Figure 1: Conceptual framework for AI prediction of low preventive practice uptake

This systematic review aims to:
(1) identify existing AI/ML models developed to predict cervical cancer preventive practice uptake based on health literacy and knowledge scores;
(2) evaluate their performance characteristics (accuracy, sensitivity, specificity, AUC);
(3) synthesize key predictors and their effect sizes; and
(4) propose a practical predictive model suitable for low resource settings.

II. METHODOLOGY

2.1 Search Strategy

A systematic review was conducted following PRISMA guidelines. Databases searched included PubMed, Scopus, Web of Science, CINAHL, and Google Scholar for publications between January 2015 and April 2025. Search terms combined MeSH terms and keywords: ("cervical cancer" OR "cervical neoplasm") AND ("preventive practice" OR "screening" OR "Pap smear" OR "HPV vaccine") AND ("health literacy" OR "knowledge" OR "awareness") AND ("machine learning" OR "artificial intelligence" OR "predictive model" OR "risk prediction" OR "random forest" OR "logistic regression"). Only peer reviewed original research articles in English were included.

2.2 Inclusion and Exclusion Criteria

Inclusion criteria: (1) adult women aged ≥18 years; (2) quantitative measurement of health literacy (validated instrument) and/or cervical cancer knowledge; (3) outcome measure of preventive practice uptake (Pap smear ever/within recommended interval, HPV vaccination initiation/completion); (4) use of predictive modelling (logistic regression, random forest, SVM, neural networks, gradient boosting, or similar); (5) report of model performance metrics (AUC, accuracy, sensitivity, specificity). Exclusion criteria: qualitative studies, case reports, editorials, conference abstracts without full data, pediatric populations, and studies without original data.

2.3 Data Extraction and Quality Assessment

Two reviewers independently extracted the following: first author, year, country, sample size, population characteristics, health literacy instrument, knowledge assessment tool, outcome definition, ML algorithm, predictor variables, performance metrics (AUC, accuracy, sensitivity, specificity, F1 score), and key findings. Quality was assessed using the PROBAST (Prediction model Risk Of Bias Assessment Tool) checklist.

2.4 Data Synthesis

Given heterogeneity in algorithms and outcome definitions, a narrative synthesis was performed. For studies reporting comparable metrics (e.g., AUC), a random effect meta-analysis was conducted for each algorithm type where at least three studies were

available. Key predictor importance scores were compiled and ranked across studies.

III. SUMMARY OF FINDINGS

3.1 Study Characteristics

Fifty-four studies met inclusion criteria, comprising 187,432 participants across 26 countries (LMICs: 38 studies; high income: 16 studies). Sample sizes ranged from 210 to 24,000. Health literacy was measured using the Newest Vital Sign (NVS; 23 studies), Short Assessment of Health Literacy (SAHL; 14 studies), or the Health Literacy Questionnaire (HLQ; 12 studies). Cervical cancer knowledge was assessed using domain specific questionnaires covering risk factors (HPV infection, smoking, multiple sexual partners), screening recommendations (starting age, interval), and HPV vaccine (target age, dosing schedule).

Preventive practice uptake outcomes varied: ever had Pap smear (35 studies), Pap smear within last 3 5 years (12 studies), HPV vaccination initiation (5 studies), and combination screening + vaccination (2 studies). Low uptake was defined as never screened or not up to date with recommendations.

3.2 Performance of Predictive AI Models

Table 1: Predictive performance of AI models for low preventive practice uptake

Algorithm	Number of studies	Mean AUC (95% CI)	Mean accuracy (%)	Mean sensitivity (%)	Mean specificity (%)
Logistic regression (traditional)	28	0.75 (0.72–0.78)	71.2	68.4	73.1
Random forest	14	0.89 (0.86–0.92)	84.5	82.1	86.3
XGBoost / Gradient boosting	9	0.91 (0.88–0.94)	86.2	84.5	87.4
Support vector machine	6	0.79 (0.74–0.83)	75.3	72.6	77.8
Neural network (shallow)	4	0.81 (0.76–0.85)	77.8	75.2	80.1
Ensemble (stacking)	3	0.92 (0.89–0.94)	87.1	85.8	88.2

Figure 2: ROC curves comparing algorithm performance (placeholder)

Ensemble methods (random forest, XGBoost, stacking) significantly outperformed traditional logistic regression ($p < 0.01$). The best performing individual model was XGBoost with an AUC of 0.94 in a large Nigerian study (n=3,420) [16]. Neural networks did not show additional benefit over gradient boosting, likely due to limited sample sizes for deep learning.

3.3 Key Predictors of Low Preventive Practice Uptake

Across all studies, the most consistent and powerful predictors (by variable importance ranking) were:

1. Functional health literacy (mean OR 3.5, 95% CI 2.8–4.2). Women with low health literacy were 3.5 times more likely to have never had a Pap smear compared to those with adequate literacy. This effect remained significant after adjusting for education and income [17,18].

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718570

2. Knowledge that HPV causes cervical cancer (mean OR 3.5, 95% CI 2.9–4.1). Women who did not know that persistent HPV infection is the primary cause of cervical cancer were substantially less likely to seek screening or accept HPV vaccination [19,20].

3. Knowledge of recommended screening interval (mean OR 2.9, 95% CI 2.4–3.5). Lack of knowledge that screening should occur every 3 5 years (depending on age and prior results) predicted never screened status.

4. Perceived susceptibility to cervical cancer (mean OR 2.3, 95% CI 1.9–2.8). Women who did not believe they were personally at risk were less likely to engage in preventive practices [21].

5. Knowledge of HPV vaccine availability/target age (mean OR 2.2, 95% CI 1.8–2.7). For vaccination outcomes, knowledge that the vaccine is most effective when given before sexual debut (ages 9 14) strongly predicted vaccine uptake among parents of adolescent girls [22].

6. Educational attainment (mean OR 2.0, 95% CI 1.6–2.5). However, when health literacy was included in the model, education added only marginal predictive value, suggesting health literacy mediates much of the education effect [23].

Figure 3: Variable importance plot from random forest model (placeholder)

3.4 Proposed Parsimonious Predictive Model

Based on the systematic review, a simplified 8 item predictive tool was derived using the top performing features from ensemble models. The tool consists of:

1. In the past 12 months, how often did you have someone help you read health information? (5 point Likert)
2. How confident are you filling out medical forms by yourself? (4 point)
3. Do you know what causes cervical cancer? (open ended: HPV, virus, infection)
4. At what age should a woman start having Pap smear screening? (21 years)
5. How often should a woman with normal results repeat Pap smear? (3 5 years)
6. Have you heard of the HPV vaccine? (Yes/No)
7. Do you believe you are at risk for cervical cancer? (Yes/No)
8. Has a doctor or nurse ever recommended cervical cancer screening to you? (Yes/No)

Using random forest with these 8 items (without demographic data) achieved AUC 0.84 (95% CI 0.81–0.87) in an external validation dataset (n=1,200), with sensitivity 84% and specificity 79% for identifying women who had never been screened [24]. Adding age and education improved AUC to 0.87, but the parsimonious version is recommended for community settings where demographic data may be sensitive or unavailable.

3.5 Deployment Considerations

Four studies evaluated implementation of AI based risk prediction in primary care and community health settings [25–28]. Key findings:

- Mobile health (mHealth) applications delivering the 8 item screener and providing immediate risk feedback were acceptable to 89% of women in rural Uganda and Nigeria.
- Community health workers (CHWs) trained to administer the screener increased targeted referrals by 2.5 fold compared to routine outreach.
- Integration with electronic medical records (EMRs) enabled automated risk flagging and recall notifications, improving screening uptake from 32% to 58% over 12 months in Kenyan pilot sites.
- Cost effectiveness analysis: AI directed targeted outreach cost 18 peradditional woman screened compared to 18 peradditional woman

screened compared to 42 per additional woman screened by universal mass media campaigns.

IV. DISCUSSION

This systematic review demonstrates that supervised machine learning models—particularly ensemble methods (random forest, XGBoost)—accurately predict low cervical cancer preventive practice uptake using brief assessments of health literacy and cervical cancer knowledge. The parsimonious 8 item model achieves clinically useful sensitivity (84%) and specificity (79%), enabling health systems to identify high risk women for targeted interventions without resource intensive comprehensive assessments.

4.1 Comparison with Traditional Approaches

Traditional risk prediction for cervical cancer screening non adherence has relied on logistic regression using demographic factors (age, education, income, marital status) [29]. These models typically achieve AUCs of 0.65–0.75 [30]. The present review shows that adding health literacy and knowledge scores substantially improves discrimination (AUC 0.85–0.92), confirming that cognitive and perceptual factors are stronger predictors than demographics alone [31,32]. Moreover, ensemble methods capture non linear interactions (e.g., low literacy with high perceived risk leading to paradoxical avoidance) that logistic regression cannot model.

4.2 Implications for Clinical Practice

Primary care clinics: The 8 item screener can be self administered on tablets in waiting rooms, with automatic risk scoring and flagging in EMRs. High risk women (probability <0.3 of being screened) trigger same day health education or navigation to a dedicated screening nurse [33].

Community health programmes: CHWs using mobile phones can administer the screener during home visits. High risk women receive tailored counselling and assistance scheduling a screening appointment or transport voucher [34].

Mass media campaigns: Instead of generic messages, AI stratification can segment women by cluster (e.g., “low literacy + low knowledge” vs. “adequate literacy + misperceptions”) and deliver culturally appropriate, literacy sensitive content via SMS or voice calls [35].

4.3 Limitations

The reviewed studies are predominantly cross sectional, meaning predictive models are trained on self reported behaviour rather than prospectively observed uptake. Only four studies reported prospective validation [36–39]. Recall bias for ever screened status is possible. Most models were developed and validated within single countries; cross cultural generalisability requires further testing. Additionally, the 8 item model has not yet been tested in diverse languages and literacy levels (e.g., pictorial versions for very low literacy).

4.4 Recommendations for Future Research

- Prospective validation: Multi site prospective cohort studies are needed to assess whether baseline prediction scores accurately predict future screening behaviour over 12 24 months.
- Intervention trials: Randomized controlled trials comparing AI directed targeted outreach versus standard education or universal screening reminders.
- Algorithm fairness: Evaluate whether models exhibit bias by age, ethnicity, or socioeconomic group; adjust thresholds to ensure equitable sensitivity.
- Implementation science: Identify barriers and facilitators to deploying AI prediction tools in low resource primary care and community settings.

V. CONCLUSION

Predictive AI models using brief health literacy and cervical cancer knowledge scores accurately identify women at high risk of low preventive practice uptake. Ensemble machine learning methods (random forest, XGBoost) outperform traditional logistic regression, achieving AUCs of 0.85–0.92. A parsimonious 8 item screener based on these models provides practical, deployable risk stratification for primary care, community health, and mobile health applications. Targeted outreach guided by AI prediction can increase screening uptake more efficiently than universal approaches. Health systems in LMICs, where cervical cancer burden is highest,

should prioritize implementation and prospective validation of these AI tools as part of comprehensive cervical cancer prevention programs.

REFERENCES

[1] Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249.

[2] Arbyn, M., Weiderpass, E., Bruni, L., et al. (2020). Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. The Lancet Global Health, 8(2), e191–e203.

[3] [3] World Health Organization. (2022). Global strategy to accelerate the elimination of cervical cancer as a public health problem. WHO Press.

[4] Schiffman, M., & Castle, P. E. (2003). The promise of global cervical cancer prevention. New England Journal of Medicine, 353(20), 2101–2104.

[5] Akinyemiju, T. F. (2012). Socio economic and health access determinants of breast and cervical cancer screening in low income countries: analysis of the World Health Survey. PLoS ONE, 7(11), e48834.

[6] Berkman, N. D., Sheridan, S. L., Donahue, K. E., et al. (2011). Health literacy interventions and outcomes: An updated systematic review. Evidence Report/Technology Assessment, 199, 1–941.

[7] Sørensen, K., Van den Broucke, S., Fullam, J., et al. (2012). Health literacy and public health: A systematic review and integration of definitions and models. BMC Public Health, 12, 80.

[8] Davis, T. C., Williams, M. V., Marin, E., et al. (2002). Health literacy and cancer communication. CA: A Cancer Journal for Clinicians, 52(3), 134–149.

[9] Tiro, J. A., Meissner, H. I., Kobrin, S., & Chollette, V. (2007). What do women in the U.S. know about human papillomavirus and cervical cancer? Cancer Epidemiology, Biomarkers & Prevention, 16(2), 288–294.

[10] Perlman, S., Wamai, R. G., Bain, P. A., et al. (2015). Knowledge and awareness of HPV vaccine and acceptability to vaccinate in sub Saharan Africa: A systematic review. PLoS ONE, 10(9), e0138213.

[11] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219.

[12] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

[13] Topol, E. J. (2019). High performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

[14] Ghaderzadeh, M., Garavand, A., & Salehnasab, C. (2025). Artificial intelligence in polycystic ovary syndrome: A systematic review of diagnostic and predictive applications. BMC Medical Informatics and Decision Making, 25(1), 427.

[15] Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.

[16] Ogunwale, A. N., Adewumi, A., & Oladapo, O. T. (2024). Predicting cervical cancer screening uptake using machine learning: A community based study in southwestern Nigeria. BMC Women's Health, 24, 215.

[17] Lindau, S. T., Tomori, C., Lyons, T., et al. (2002). The association of health literacy with cervical cancer prevention knowledge and Pap smear adherence. Journal of Women's Health, 11(5), 449–456.

[18] White, S., Chen, J., & Atchison, R. (2008). Relationship of preventive health practices and health literacy: a national study. American Journal of Health Behavior, 32(3), 227–242.

[19] Denny Smith, T., Bair, B., & Clifton, S. (2019). Cervical cancer knowledge and screening uptake among African American women. Journal of Racial and Ethnic Health Disparities, 6(4), 779–787.

[20] Wong, L. P., Wong, P. F., & Abubakar, S. (2020). HPV vaccine knowledge and acceptability among Malaysian women. Vaccine, 38(8), 1923–1930.

[21] Brewer, N. T., Chapman, G. B., Gibbons, F. X., et al. (2007). Meta analysis of the relationship between risk perception and health behavior. Health Psychology, 26(2), 136–145.

[22] Dochez, C., Burnett, R. J., & Mbatha, S. (2019). Knowledge of HPV and HPV vaccine among adolescent girls and young women in South Africa. South African Medical Journal, 109(5), 339–344.

[23] Bennett, I. M., Chen, J., Soroui, J. S., & White, S. (2009). The contribution of health literacy to disparities in self rated health status and preventive health behaviors in older adults. Annals of Family Medicine, 7(3), 204–211.

[24] Nwankwo, K. C., Aniebue, U. U., & Onyeka, T. C. (2025). External validation of an 8 item machine learning screener for cervical cancer screening non adherence in rural Nigeria. International Journal of Gynecology & Obstetrics, 168(2), 345–352.

[25] Huchko, M. J., Sneden, J., Zakaras, J. M., et al. (2018). A mobile health intervention for cervical cancer screening in Kenya. JCO Global Oncology, 4, 1–9.

[26] Asgary, R., Adedoyin, A. C., Naderi, R., et al. (2020). Community based mobile health intervention for cervical cancer screening in Nigeria. BMJ Global Health, 5(2), e001978.

[27] Tran, D. N., & Gibson, L. (2021). Implementation of electronic medical record based risk prediction for cervical cancer screening. Journal of the American Medical Informatics Association, 28(5), 987–995.

[28] Oluoch, T., Mwangi, C., & Kariuki, J. (2024). Cost effectiveness of AI directed outreach versus mass media for cervical cancer screening in Kenya. Health Policy and Planning, 39(3), 256–267.

[29] Sabatino, S. A., White, M. C., Thompson, T. D., & Klabunde, C. N. (2015). Cancer screening test use among U.S. women. American Journal of Preventive Medicine, 48(4), 386–398.

[30] Nelson, H. D., Pappas, M., Cantor, A., et al. (2019). Risk assessment for cervical cancer screening: a systematic review. Annals of Internal Medicine, 171(1), 29–38.

[31] Bazargan, M., Lucas Wright, A., Jones, L., et al. (2017). Health literacy and cancer screening among African American women. Journal of Cancer Education, 32(3), 527–534.

[32] Sentell, T., Braun, K. L., Davis, J., & Davis, T. (2015). Health literacy and meeting breast and cervical cancer screening guidelines among Asians and whites in California. SpringerPlus, 4, 432.

[33] Zapka, J. G., Taplin, S. H., Solberg, L. I., & Manos, M. M. (2003). A framework for improving the quality of cancer screening: the case of cervical cancer screening. Cancer Epidemiology, Biomarkers & Prevention, 12(4), 303–312.

[34] Lewin, S., Munabi Babigumira, S., Glenton, C., et al. (2010). Lay health workers in primary and community health care for maternal and child health and the management of infectious diseases. Cochrane Database of Systematic Reviews, 3, CD004015.

[35] Kreuter, M. W., & Wray, R. J. (2003). Tailored and targeted health communication: strategies for enhancing information relevance. American Journal of Health Behavior, 27(Suppl 3), S227–S232.

[36] Nygård, M., & Kim, E. (2023). Prospective validation of a machine learning model for cervical cancer screening non adherence. Preventive Medicine, 170, 107468.

[37] Aerts, J., & Van den Bergh, R. (2023). Predicting cervical cancer screening participation using health literacy and knowledge: a prospective cohort study. European Journal of Public Health, 33(Suppl 2), ckad160.1367.

[38] Okonofua, F. E., & Ntoimo, L. F. (2024). Prospective validation of an AI based cervical cancer screening risk score in Nigerian women. African Journal of Reproductive Health, 28(5), 45–56.

[39] Sankaranarayanan, R., & Basu, P. (2025). Implementation of AI driven risk stratification for cervical cancer prevention: a multi country prospective study. The Lancet Digital Health, 7(2), e89–e101.