



Contribution ID: 14

Type: **not specified**

An Ensemble Machine-Learning based Approach to predict Cervical High-risk Human Papillomavirus among women in Ibadan, Nigeria

Tuesday, 22 July 2025 12:15 (15 minutes)

Persistence of cervical high-risk human papillomavirus (hrHPV) is a necessary cause of cervical cancer (CC) that remains a significant public health concern globally. Although CC is largely preventable, it is still cause of mortality in adult women especially in sub-Saharan Africa. Screening for CC precancer and early invasive cancer is pivotal to a successful elimination strategy in any country. This study provides insight on how to efficiently profile women with cervical hrHPV by using an Ensemble Machine Learning (EML) classifier.

This analysis used data from Sexual Behaviours and HPV Infections among Nigerians in Ibadan (SHINI) study to develop models for cervical hrHPV. Relevant data were extracted to develop an ensemble model. The ensemble model was based on Logistic Regression (LR), Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). The data was divided into training (70) and testing sets (30). Model performance was assessed using Area Under the Receiver Operating Curve (AUC-ROC), accuracy, and F1-score. A value greater than or equal to 0.7 was adjudged as a good model.

The features extracted included age, ethnicity, income, multiple sexual partners, condom use, alcohol use, cigarette smoking, illicit drug use, knowledge of HPV, ever had anal sex, and prior anal HPV infection. The AUC for training (testing) data was 0.74(0.83) for EML; 0.77(0.79) for LR, 0.99(0.63) for DT, 0.73(0.78) for NB, 0.89(0.73) for KNN, 0.77(0.78) for SVM, and 0.73(0.74) for ANN. Accuracy for training (testing) data was 0.73(0.76) for EML, 0.72(0.71) for LR, 0.98(0.63) for DT, 0.66(0.67) for NB, 0.83(0.70) for KNN, 0.73(0.69) for SVM, and 0.73(0.73) for ANN. F1-score were 0.78(0.79) for EML, 0.79(0.76) for LR, 0.99(0.66) for DT, 0.71(0.68) for NB, 0.85(0.72) for KNN, 0.78(0.73) for SVM, and 0.79(0.76) for ANN in training (testing) respectively.

The EML model demonstrated superior predictive performance for cervical hrHPV, highlighting its potential to enhance risk stratification and inform targeted screening and intervention strategies in Nigeria and other resource-limited settings.

Primary author: Mrs OKE, Deborah (Department of Epidemiology & Medical Statistics, Faculty of Public Health, University of Ibadan, Ibadan, Nigeria. HPV Consortium, College of Medicine, University of Ibadan, Ibadan, Nigeria.)

Co-authors: Prof. WATSON-JONES, Deborah (1. Department of Clinical Research, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. 2. Mwanza Intervention Trials Unit, National Institute of Medical Research Mwanza, Tanzania); Prof. MORHASON-BELLO, Imran (1. HPV Consortium, College of Medicine, University of Ibadan, Ibadan, Nigeria. 2. Department of Obstetrics & Gynaecology, Faculty of Clinical Sciences, University of Ibadan, Nigeria. 3. Institute for Advanced Medical Research and Training, College of Medicine, University of Ibadan, Ibadan, Nigeria); Prof. ADEWOLE, Isaac (Department of Obstetrics & Gynaecology, Faculty of Clinical Sciences, University of Ibadan); Prof. AKINYEMI, Joshua (1. Department of Epidemiology & Medical Statistics, Faculty of Public Health, University of Ibadan, Ibadan, Nigeria. 2. Institute for Advanced Medical Research and Training, College of Medicine, University of Ibadan, Ibadan, Nigeria. 3. Infectious Disease Institute, College of Medicine, University of Ibadan, Ibadan, Nigeria.); Prof. HOU, Lifang (Department of Preventive

Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA); Dr ZHENG, Yinan (Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA)

Presenter: Mrs OKE, Deborah (Department of Epidemiology & Medical Statistics, Faculty of Public Health, University of Ibadan, Ibadan, Nigeria. HPV Consortium, College of Medicine, University of Ibadan, Ibadan, Nigeria.)

Session Classification: Contributed talk: Room-2 (Statistics, Biostatistics and Epidemiology)

Track Classification: Sciences: Biostatistics/Medical Statistics