Research Square

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

Machine Learning for Predicting Childhood Immunization Status and Identifying key Predictors Among Children Aged 12–23 Months in Sub-Saharan Africa: A Multi-Country DHS Analysis

Makda Fekadie Tewelgne

makdafekadeal@gmail.com

University of Gondar

Selamawite Fekadie Tewoligne

Debre Tabour University

Dejen Kahsay Asgedom

Samara University

Degsew Ewunetie Anteneh

Woldia University

Mulualem Endeshaw Zeleke

University of Gondar

Alemken Eyayu Abuhay

University of Gondar Comprehensive Specialized Referral Hospital

Berihun Agegn Mengistie

University of Gondar

Haymanot Nega Alemu

University of Gondar

Kindu Yinges Wondie

University of Gondar

Tazeb Alemu Anteneh

University of Gondar

Amare Belete Getahun

University of Gondar

Jenberu Mekurianew Kelkay

Debark University

Halima Ayalew Kebede

Woldia University

Abraham Keffale Mengistu

Debre Markos University

Andualem Enyew Gedefaw
University of Gondar
Tadele Chekol Maru
University of Gondar
Fentahun Bikale Kebede
Federal Ministry of Health

Research Article

Keywords: Childhood immunization, Machine learning, Explainable artificial intelligence, SHAP, Demographic and Health Survey, Sub-Saharan Africa, Predictive modeling

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11348273/v1

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

Additional Declarations: No competing interests reported.

Machine Learning for Predicting Childhood Immunization Status and Identifying key Predictors Among Children Aged 12–23 Months in Sub-Saharan Africa: A Multi-Country DHS Analysis

Makda Fekadie Tewelgne[1*],Selamawite Fekadie Tewoligne[2],Dejen Kahsay Asgedom[3], Degsew Ewunetie Anteneh[4],Mulualem Endeshaw Zeleke[5] , Alemken Eyayu Abuhay[6], Berihun Agegn Mengistie[7],Haymanot Nega Alemu[8], Kindu Yinges Wondie[8], Tazeb Alemu Anteneh[8], Amare Belete Getahun[9],Jenberu Mekurianew Kelkay[10] , Halima Ayalew Kebede[11], Abraham Keffale Mengistu [12], Andualem Enyew Gedefaw [1], Tadele Chekol Maru[1], Fentahun Bikale Kebede[13]

[1]Department Health informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[2]Department of Neonatal, College of Medicine and Health Sciences, Debre Tabour University, Debre Tabour, Ethiopia

[3]Department of Public Health, College of Medicine and Health Science, Samara University, Samara, Ethiopia

[4]Department of Clinical Midwifery, School of Midwifery, College of Medicine and Health Science, Woldia University, Woldia, Ethiopia

[5]Department of Anesthesia, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[6]Department of Clinical Midwifery, College of Medicine and Health Sciences, University of Gondar Comprehensive Specialized Referral Hospital, Gondar, Ethiopia

[7]Department of General Midwifery, School of Midwifery, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[8]Department of Clinical Midwifery, School of Midwifery, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[9]Department of Anesthesia, School of Medicine, College of Medicine and Health sciences, University of Gondar, Gondar, Ethiopia

[10]Department of Public Health, College of Health Sciences, Debark University, Debark, Ethiopia

[11]Department of Public Health, College of Health Sciences, Woldia University, Woldia, Ethiopia.

[12]Department of Health Informatics, College of Medicine and Health Sciences, Debre Markos University, Debre Markos, Ethiopia

[13] Strategic Affairs Executive Office, Federal Ministry of Health, Addis Ababa, Ethiopia

*Corresponding author

Makda Fekadie Tewelgne

Email: makdafekadeal@gmail.com

Email-
makdafekadeal@gmail.com,selamewitefekadie@gmail.com,dejenk2@gmail.com,degsewewunetie@gmail.com,muleaman45@gmail.com,alexeyayu6@gmail.com ,berihunagegn21@gmail.com, haymnigat@gmail.com,kinduyinges2010@gmail.com,tazebalemu05@gmail.com,amarebelete2218@gmail.com,jenberu.uog11@gmail.com,halimaayalew0977@gmail.com,keffaleabrahame2@gmail.com,andualemenyew@gmail.com,tadelle2007@gmail.com, fentahunbikale09@gmail.com

Abstract

Background:

Childhood immunization is one of the most cost-effective public health interventions for preventing vaccine-preventable morbidity and mortality; yet substantial coverage gaps persist across Sub-Saharan Africa. These gaps are driven by persistent socioeconomic inequities, geographical disparities, and barriers to the provision of consistent service delivery that impact immunization uptake across multiple levels of influence. Machine learning possesses the capability to identify high-risk children and the key factors leading to under-immunization by examining complex relationships within large population datasets. These studies aimed to development and evaluate standard Machine learning models for predicting of childhood immunization status and identifying key important predictors among children aged 12-23 months in Sub-Saharan Africa using multi-country DHS data.

Methods: A pooled cross-sectional analysis was performed utilizing DHS Kids Recode datasets from 22 Sub-Saharan African (SSA) nations, collected between 2019 and 2025, encompassing a total of 49,817 children aged 12 to 23 months. Seven machine learning algorithms were developed and assessed, including Logistic Regression, Decision Tree, Random Forest, K-

Nearest Neighbors, Artificial Neural Network, XGBoost, and LightGBM, employing metrics such as ROC-AUC, PR-AUC, accuracy, precision, recall, F1-score, and calibration. To interpret model predictions and pinpoint significant factors influencing childhood immunization, SHapley Additive exPlanations (SHAP) were utilized.

Results: Among the machine learning models evaluated, the Random Forest algorithm exhibited the most superior predictive capability regarding childhood immunization status, attaining an accuracy of 94.24%, alongside a ROC-AUC of 0.9875 and a PR-AUC of 0.9890. The prevalence of complete immunization among children aged 12 to 23 months was found to be 40.11%.

The interpretation of the model using SHAP highlighted the key factors that contribute to the predictions of childhood immunization status. The most significant predictors included adequate antenatal care (ANC) utilization (four or more visits), institutional delivery, postnatal care (PNC) receipt, maternal education (secondary and higher education), partner education (secondary and higher education), older maternal age, skilled birth attendance, and urban residency. These factors collectively emphasize the critical role of maternal healthcare engagement, educational empowerment, and equitable access to essential health services throughout the continuum of maternal and child care in influencing childhood immunization status.

Conclusion: Machine learning models have exhibited robust capabilities in predicting childhood immunization status and in identifying significant predictors of immunization outcomes. The integration of explainable artificial intelligence boosts the interpretative aspect of these models, thereby facilitating the development of targeted, data-driven strategies that aim to strengthen vaccination programs and improve childhood immunization rates in sub-Saharan Africa. Such

advancements are crucial for optimizing public health initiatives and ensuring better health for children in the region.

Keywords: Childhood immunization; Machine learning; Explainable artificial intelligence; SHAP; Demographic and Health Survey; Sub-Saharan Africa; Predictive modeling.

Introduction

Childhood immunization serves as a crucial aspect of public health, offering prompt protection against diseases that can be prevented by vaccines, while also decreasing related morbidity and mortality through a method that is effective, equitable, and cost-efficient (1-3). In recent years, the expansion of routine immunization programs has resulted in substantial reductions in childhood deaths caused by diseases such as measles, pertussis, diphtheria, poliomyelitis, and pneumococcal infections, representing a significant achievement in global initiatives aimed at improving child survival (4, 5).

The World Health Organization (WHO) advocates for a fundamental immunization schedule for infants, which generally comprises Bacillus Calmette–Guérin (BCG), oral and/or inactivated poliovirus vaccines, diphtheria-pertussis-tetanus (DPT) vaccines, measles-containing vaccine, pneumococcal conjugate vaccine (PCV), and rotavirus vaccine, aimed at ensuring prompt and comprehensive protection during the most vulnerable phase of life (6, 7). Despite considerable global progress, significant disparities in immunization remain widespread, particularly in low- and middle-income countries (LMICs), where health system limitations, geographical barriers, sociocultural factors, and the unequal distribution of health resources continue to obstruct access to comprehensive vaccination services (8, 9).

Globally, millions of children still miss essential routine vaccines each year—particularly in 2022 when an estimated 14 million were “zero-dose” with no routine immunizations, and many others were only partially vaccinated (10). This issue is particularly pronounced in low- and middle-income countries (LMICs), hindering advancements toward the Immunization Agenda 2030 (IA2030), which aims for universal and equitable access to life-saving vaccines and the mitigation of immunization disparities (11, 12). The inability to achieve comprehensive coverage heightens the risk of outbreaks of vaccine-preventable diseases and imposes further pressure on already vulnerable health systems (13).

Sub-Saharan Africa (SSA) bears a disproportionate share of the global burden of vaccines-preventable diseases, and despite the expansion of immunization programs across the region, vaccination rates still exhibit considerable variation both between countries and within individual nations (14, 15). In many Sub-Saharan African contexts, the rate of full immunization coverage is frequently below the global target of 90%, often fluctuating between 50% and 70%, and in some areas, it can drop to as low as 40% to 50% for children aged 12 to 23 months (16, 17). These disparities underscore the persistent inequalities in both the accessibility and utilization of vaccination services, which are shaped by systemic challenges including insufficient health infrastructure, a shortage of trained professionals, geographical obstacles, and vulnerable vaccine supply chains (18, 19). Thus, children in Sub-Saharan Africa encounter a greater likelihood of preventable infections and suffer from significantly elevated mortality rates in contrast to their counterparts in affluent regions (20, 21). These enduring disparities endanger the achievement of immunization objectives at both the regional and global levels, underscoring the importance of implementing focused, evidence-based strategies (17).

The repercussions of insufficient immunization reach beyond the immediate threat of infection, as children who are either unvaccinated or inadequately vaccinated face a heightened vulnerability to recurrent illnesses, malnutrition, cognitive development issues, and long-term disabilities (22, 23). On a broader scale, low vaccination coverage undermines herd immunity, elevates the risk of outbreaks, and incurs significant economic burdens on families and healthcare systems (24, 25). Thus, enhancing immunization coverage is crucial not only for the survival of children but also for promoting health equity and fostering sustainable development (26).

A substantial body of evidence identifies childhood immunization determinants operating at individual, household, and community levels, including maternal factors such as low educational attainment, limited health knowledge, unemployment, and reduced media exposure, health service utilization factors such as inadequate antenatal care (ANC), lack of postnatal care (PNC), home delivery, and absence of health insurance, as well as household and contextual factors including poverty, high birth order, rural residence, distance to health facilities, and regional disparities in healthcare access—also play significant roles (20, 27-30). The interplay of these determinants is often complex and non-linear, complicating the ability to fully understand their collective impact through traditional analytical methods (31).

Traditional regression-based models continue to be valuable for inference and hypothesis testing; however, they often face limitations when addressing high-dimensional data, non-linear relationships, and intricate interactions among predictors (32, 33). Traditional regression-based models continue to be valuable for inference and hypothesis testing; however, they often face limitations when addressing high-dimensional data, non-linear relationships, and intricate interactions among predictors (34).

Machine learning (ML) serves as an additional technique for predicting childhood immunization status, where explainable methods such as SHAP facilitating interpretability by underscoring the most critical variables (35, 36).

Despite the increasing application of ML in health research, its use for predicting childhood immunization status in sub-Saharan Africa is still relatively scarce. The existing literature predominantly relies on analyses from single countries (37-44) , East Arica (45-48) , and Sub-Saharan Africa (20, 49) or conventional statistical methods, which may not fully reflect the complex socioeconomic, health-system, and geographic heterogeneity across the region. Furthermore, the limited availability of multi-country ML studies (50-53), highlight the potential of scalable and interpretable predictive approaches for identifying childhood immunization status using population-level data. The availability of harmonized, nationally representative DHS data therefore provides an important opportunity to develop predictive models capable of capturing determinants of immunization status across diverse SSA populations.

Consequently, this study employed machine learning to predict childhood immunization status and identify key factors among children aged 12–23 months in SSA using multi-country DHS data collected between 2019 and 2025. The study seeks to provide interpretable evidence that can inform targeted interventions and reinforce childhood immunization efforts throughout the region.

Methods

Study Design and Setting

A community-based cross-sectional study was implemented using pooled, nationally representative datasets from the Demographic and Health Survey (DHS) across 22 Sub-Saharan African countries, conducted between 2019 and 2025.

The DHS program employs standardized and internationally comparable survey methodologies to produce extensive population-based evidence regarding maternal and child health, healthcare access and utilization, immunization coverage, as well as significant demographic, socioeconomic, and behavioral health determinants. The latest DHS datasets from each participating country were harmonized and compiled into a consistent analytical dataset, assisting a thorough assessment of childhood immunization status and its related factors for children aged 12 to 23 months.

Study Population and Data Sources

We analyzed pooled Demographic and Health Survey (DHS) Kids Recode (KR) datasets to predict childhood immunization status and identify its associated factors among children aged 12–23 months in Sub-Saharan Africa. The DHS KR datasets provide detailed child-level information, including vaccination history, child characteristics, maternal characteristics, and household factors. The pooled dataset included data from 22 Sub-Saharan African countries: Angola, Burkina Faso, the Democratic Republic of the Congo, Côte d'Ivoire, Ethiopia, Gabon, Ghana, The Gambia, Kenya, Liberia, Lesotho, Madagascar, Mali, Mauritania, Malawi, Mozambique, Nigeria, Rwanda, Sierra Leone, Senegal, Tanzania, and Zambia.

The DHS uses a stratified two-stage cluster sampling design. In the first stage, enumeration areas (EAs) are selected, and households are systematically selected within each EA (48); Child-level

data were obtained from the DHS KR datasets, which include information collected through standardized household surveys.

To obtain nationally representative estimates for descriptive and prevalence analyses, DHS sampling weights (v005), stratification (v022), and clustering (v021) were applied according to DHS analytical guidelines. Sampling weights were normalized by dividing by 1,000,000 as recommended by DHS procedures (54).

For predictive modeling, pooled individual-level data from Sub-Saharan African countries were used. Sampling weights were not incorporated during model training because the objective of the analysis was individual-level prediction rather than population-level inference. Machine learning algorithms are designed to learn predictive patterns from observed data distributions; therefore, each observation was treated equally during model development to optimize predictive performance (55).

Although DHS data are collected using a complex survey design, incorporating survey weights, clustering, and stratification directly into standard machine learning algorithms remains methodologically challenging (56). Therefore, these design features were not included during model training. However, they were fully considered in descriptive and prevalence analyses to ensure nationally representative estimates.

Country identifiers were retained for data traceability and exploratory analysis but were not included as predictors in the machine learning models. No hierarchical (fixed or random effects) structure was incorporated. Between-country variation in childhood immunization status and predictor distributions was examined descriptively before model development to assess heterogeneity across countries.

This unified modeling approach was employed to develop a unified predictive framework designed to assess childhood immunization status across Sub-Saharan Africa. Though pooling data improves generalizability and enables identifying common factors, it may lessen sensitivity to differences in vaccination rates, health system characteristics, and contextual factors unique to each nation. No calibration or adjustment for the prevalence of outcomes specific to individual countries was performed; consequently, model predictions should be interpreted primarily as indicators of overall predictive performance rather than as conclusive country-specific risk estimates.

The pooled DHS datasets initially included children from the selected Sub-Saharan African countries. After applying eligibility criteria, restricting the analysis to children aged 12–23 months, performing data cleaning, handling missing values, and excluding incomplete records, the final analytic sample consisted of 49,817 children.

Inclusion and Exclusion Criteria

Sub-Saharan African nations that have Demographic and Health Survey (DHS) Kids Recode (KR) datasets available from 2019 to 2025 were eligible for inclusion. Excluded were nations that did not have DHS data available within the designated time frame. Children between the ages of 12 and 23 months who had access to data on their childhood immunization status and pertinent predictor characteristics made up the analysis group. Only children born within the five years prior to each survey were included in the KR datasets in accordance with DHS data collection protocols.

The final analysis did not include children with implausible values, inconsistent vaccination responses, or lacking information regarding the immunization outcome. Prior to model building, missing values in predictor variables were filled in during the data pretreatment phase using suitable imputation techniques.

Study Variables

The primary outcome of interest was the immunization status of children aged 12–23 months. In line with the DHS-8 guidelines and national immunization schedules (54, 57), immunization status was defined based on the Expanded Program on Immunization (EPI) recommendations (58), and vaccination information available in the DHS Kids Recode (KR) datasets. The assessment included Bacille Calmette–Guérin (BCG), polio vaccine, pentavalent vaccine (DPT-HepB-Hib), pneumococcal conjugate vaccine (PCV), rotavirus vaccine, and measles-containing vaccine. Children were classified as fully immunized if they had received all recommended doses of the vaccines included in this study according to their respective national immunization schedules; otherwise, they were classified as not fully immunized. The outcome variable was coded as a binary variable, where 0 indicated not fully immunized and 1 indicated fully immunized.

Independent variables: The predictors included child-related, maternal-related, maternal healthcare utilization, household socioeconomic, environmental, and information-related characteristics. Child-related characteristics included child's sex, child's age, multiple birth

status, birth order, birth interval, birth size, current breastfeeding status, diarrhea in the last two weeks, fever in the last two weeks, and acute respiratory infection (ARI) symptoms.

Maternal-related characteristics included maternal age, maternal education level, maternal employment status, marital status, age at first birth, parity, and maternal literacy status.Maternal healthcare utilization characteristics included antenatal care (ANC) visits, skilled birth attendance, place of delivery, postnatal care (PNC) receipt, health insurance coverage, and perceived distance to health facility as a barrier. Household socioeconomic, environmental, and information-related characteristics included household wealth index, place of residence, household size, sex of household head, religion, partner's education level, source of drinking water, toilet facility, electricity access, cooking fuel type, mass media exposure, and internet use.

All variables were defined and coded according to standard DHS definitions and recode manuals to ensure clarity, consistency, and reproducibility (54).

Operational Definition

Childhood immunization status: The outcome variable was defined based on the Expanded Program on Immunization (EPI) recommendations, DHS-8 guidelines, and national immunization schedules. Children aged 12–23 months were classified as fully immunized if they had received all recommended vaccine doses included in this study, including Bacille Calmette–Guérin (BCG), polio vaccine, pentavalent vaccine (DPT-HepB-Hib), pneumococcal conjugate vaccine (PCV), rotavirus vaccine, and measles-containing vaccine, according to the respective national immunization schedules. Children who had missed at least one recommended vaccine

dose were categorized as not fully immunized. The outcome variable was coded as a binary variable: 0 = not fully immunized and 1 = fully immunized.

Fully immunized child: A child aged 12–23 months who received all recommended vaccine doses included in the study according to the national immunization schedule, based on DHS vaccination records or maternal recall.

Not fully immunized child: A child aged 12–23 months who missed at least one recommended vaccine dose included in the study.

Ethical Considerations and Data Access

Access to the Demographic and Health Survey (DHS) datasets was obtained through the DHS Program website (www.dhsprogram.com) following submission and approval of a data access request describing the study objectives and intended use of the datasets. The datasets utilized in this study were fully anonymized and did not contain any direct personal identifiers; thus, participant privacy and confidentiality were maintained throughout data management, analysis, and reporting.

The original DHS surveys were executed in accordance with internationally acknowledged ethical standards, which included informed consent, voluntary participation, confidentiality protections, and the safeguarding of respondents' rights. Given that this research involved a secondary analysis of publicly available, de-identified DHS datasets, no additional ethical approval was needed. All activities related to data handling, analysis, and dissemination were conducted in compliance with DHS data use policies and ethical guidelines, ensuring that no individual participant could be identified from the results of the study.

Data Pre-Processing

Data Integration and Harmonization

Prior to model development, individual-level data were pooled from selected Sub-Saharan African countries using the Demographic and Health Surveys (DHS) Kids Recode (KR) datasets. The integration process involved a rigorous harmonization of variables across survey years and countries to ensure consistency in variable names, coding schemes, and operational definitions. All variables were standardized in accordance with DHS coding conventions and study-specific definitions to produce a comparable pooled dataset for childhood immunization prediction analysis (54).

The harmonization procedure was vital to ensure that the observed differences in childhood immunization coverage and predictor distributions among countries truly reflected real contextual variations, rather than inconsistencies in the construction, coding, or measurement of variables across DHS surveys. To support comparability between countries, a structured harmonization framework was applied prior to the pooling of datasets. This framework involved mapping country-specific variables into a cohesive analytical structure, recoding categorical variables into standardized categories, and aligning variable definitions across the DHS KR datasets. In cases where response categories differed across surveys, conceptually equivalent categories were combined to preserve interpretability and consistency across countries.

To minimize measurement heterogeneity and improve the validity of pooled analyses, only variables with comparable definitions and consistent availability across countries were retained. Rule-based data cleaning procedures were also applied, including removal of duplicate records,

correction of inconsistent coding, and exclusion of invalid or implausible values based on DHS data quality standards (54).

Country identifiers were retained only for descriptive analyses to assess cross-country variations in childhood immunization status and population characteristics. DHS sampling weights were applied in descriptive analyses to account for the complex survey design and generate nationally representative estimates. The final harmonized pooled dataset was subsequently used for machine learning model development and evaluation.

Data Splitting

The dataset was divided using a stratified hold-out method to ensure reliable performance evaluation and to mitigate overfitting (43). The entire analytical sample included 49,817 children aged 12 to 23 months, which was randomly divided into a training set of 39,853 children (80%) and a testing set of 9,964 children (20%). Stratification was implemented to maintain the existing distribution of immunization status in both subsets. The training data were employed for preprocessing, feature selection, resampling, and model development, while the testing data were preserved solely for the final evaluation to prevent data leakage and ensure an unbiased assessment of predictive performance.

Data Cleaning

Data cleaning was conducted to ensure the accuracy, consistency, and reliability of the dataset (59). In line with DHS analytical guidelines (54), flagged observations were reviewed to identify inconsistencies and implausible values, thereby improving the quality of the analytic dataset prior to model development.

Missing data were minimal across all variables, with no variable exceeding 2.1% missingness. This indicated a high degree of completeness and a low likelihood of bias due to missing information. To address incomplete observations, K-nearest neighbors imputation was applied using k=5 and Euclidean distance as the similarity metric (60). The use of k=5 was informed by prior evidence suggesting that values between 3 and 10 offer a reasonable balance between bias and variance in imputation performance (61).

To prevent data leakage, imputation was performed only within the training data during cross-validation, and the fitted parameters were applied to the test set (62). Although KNN imputation may introduce limited bias in heterogeneous datasets, it was selected for its ability to preserve multivariate relationships and perform well in structured survey data (63, 64).

To address anomalous observations, the Isolation Forest algorithm was used to detect outliers (65), and observations with substantial influence on model stability were assessed through sensitivity analysis and removed where necessary.

All preprocessing steps were carried out within the machine learning pipeline on the training data only, while the test set remained unseen and was reserved for final evaluation.

Data Transformation

Data transformation was undertaken to prepare the harmonized DHS variables for machine learning analysis. Continuous variables were standardized using z-score normalization to ensure comparability across predictors and to enhance model stability during training. Categorical variables were converted into binary indicators using one-hot encoding, thereby allowing non-numeric variables to be processed without introducing artificial ordinal structures (66).

Data Discretization

Continuous variables were purposefully discretized into clinically relevant categories to improve interpretability, lessen the effect of extreme values, and facilitate subsequent analyses (67). The framework for categorization was informed by both clinical relevance and statistical factors, adhering to the guidelines set forth by the Demographic and Health Survey (DHS) (54). For instance, maternal age was classified into 15-19 years, 20-34 years, and 35-49 years, while maternal education was categorized as no education, primary, secondary, and higher (54). This approach maintained analytical rigor while producing variables that were more suitable for subgroup analyses and policy-oriented interpretation.

Feature Selection

To enhance model parsimony and predictive accuracy, the most informative factors were identified using Recursive Feature Elimination (RFE) (68).

This procedure was applied exclusively to the training data within cross-validation to ensure that feature selection remained independent of the test set and to minimize the risk of data leakage (69). Predictors were eliminated iteratively according to their contribution to model performance, and the resulting subset was retained as the most predictive set of features. This approach reduced redundancy, improved computational efficiency, and strengthened the interpretability of the final model while preserving a rigorous and leakage-safe selection process.

Data Analysis Procedure

Data analysis was performed using Stata version 17 and Python version 3.12.13 for descriptive analysis, data preprocessing, machine learning model development, evaluation, and interpretation of pooled 2019–2025 Demographic and Health Survey (DHS) datasets from Sub-

Saharan African countries. Python-based analyses were conducted using relevant libraries, including scikit-learn (1.6.1), pandas (2.2.2), NumPy (2.0.2), XGBoost (3.2.0), SHAP (0.51.0), imbalanced-learn, and Matplotlib. Fixed package versions and controlled random seeds were used to ensure reproducibility of the analytical workflow.

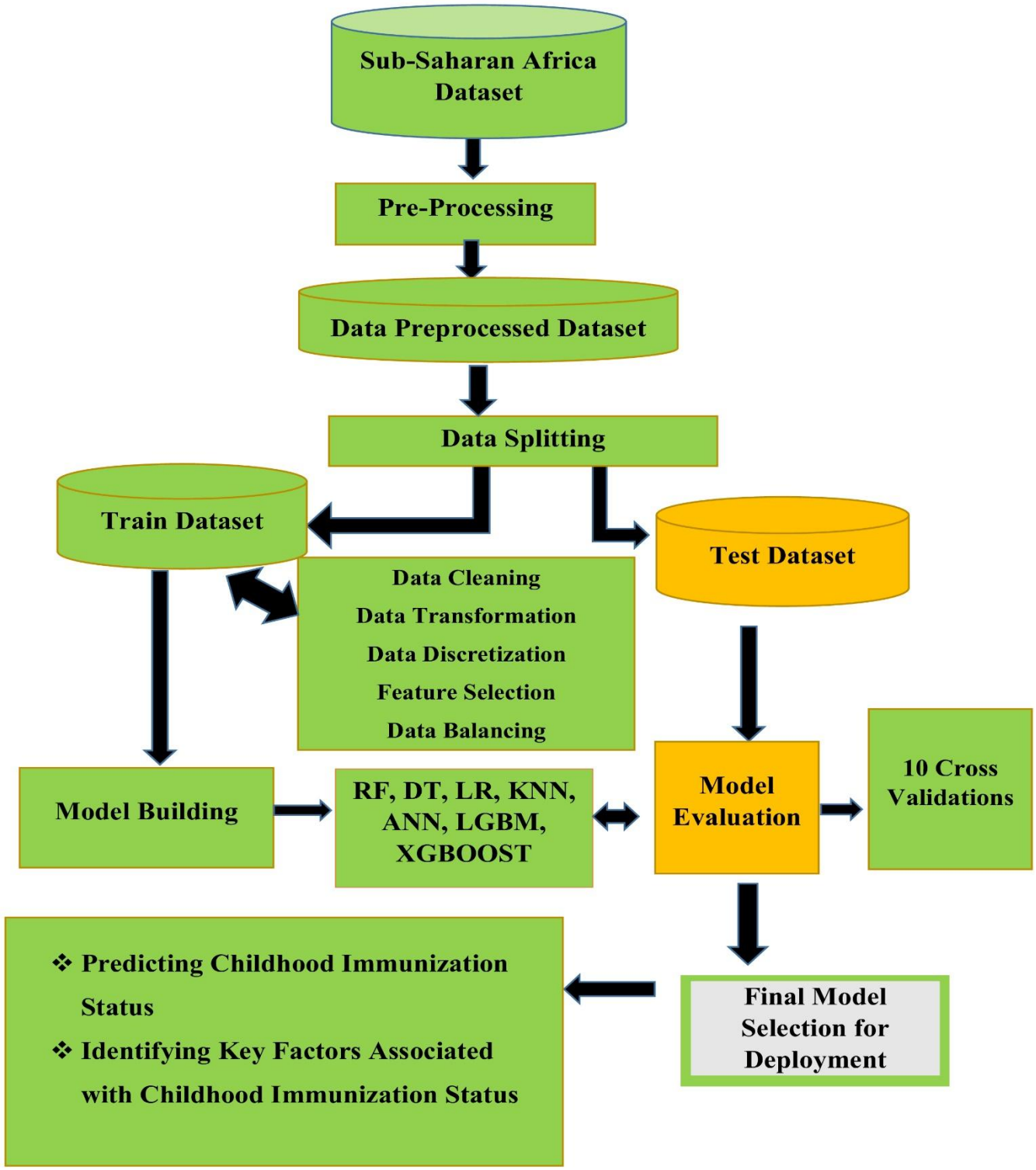
Descriptive analyses were conducted using DHS sampling weights to account for the complex survey design and generate representative estimates of childhood immunization status and predictor characteristics. Machine learning analyses were performed using unweighted data to optimize predictive modeling. Supervised machine learning algorithms were applied due to their ability to capture complex, nonlinear relationships and interactions among multiple child, maternal, healthcare, and household-level characteristics.

The analytical workflow included data preprocessing, feature selection, class imbalance handling, model training, hyperparameter optimization, performance evaluation, and model interpretation. Multiple supervised algorithms, including Random Forest, Decision Tree, Logistic Regression, K-Nearest Neighbor, Artificial Neural Network, LightGBM, and XGBoost, were developed to predict childhood immunization status. The overall analytical workflow is presented in Figure 1.

In order to address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was implemented solely on the training dataset within the cross-validation framework to avoid information leakage. All preprocessing procedures, which encompassed imputation, encoding, scaling, feature selection, and resampling, were incorporated into a machine learning pipeline that is free from leakage, where transformations were fitted only on training folds and then applied to validation or test datasets.

During the model development phase, Recursive Feature Elimination (RFE) was employed to pinpoint the most significant predictor variables. The evaluation of model performance was conducted using a variety of metrics, including accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, and calibration measures.

In order to improve interpretability, SHAP (Shapley Additive Explanations) was utilized following the training of the model to measure the contributions of predictors and elucidate the decisions made by the model, thus identifying the essential factors affecting childhood immunization status.



414

415 **Figure 1 Work flow of study**

The figure elucidates the main steps of the study, from DHS data selection and preprocessing to data splitting, model development, evaluation, prediction of childhood immunization status, and identification of key predictors.

Class Balancing

To reduce the detrimental effects of outcome imbalance on model estimate, a resampling strategy was employed to prevent the learner from being unduly influenced by the majority class (70).Specifically, the Synthetic Minority Over-Sampling Technique (SMOTE) was applied exclusively to the training data within cross-validation folds, whereby synthetic minority-class observations were generated through interpolation between nearest neighboring cases, thereby preserving local data structure while enriching representation of the minority class (71, 72). A neighborhood size of k = *5* was adopted, with Euclidean distance used as the basis for constructing synthetic samples.

In contrast to naive random oversampling, SMOTE offers the advantage of reducing overfitting by generating new, credible observations rather than simply replicating existing data points, thereby enhancing the model's capacity to form more resilient decision boundaries and elevating classification performance (73).

Given the heterogeneous data structure in the dataset, categorical variables were encoded prior to the resampling process to ensure that the creation of synthetic data occurred within a valid numerical feature space.

The test set was completely excluded from any preprocessing and was exclusively used for the final evaluation of the model. To prevent the risk of information leakage, the dataset was split into training and test subsets before any analytical transformations were executed.

.

Model Selection and Development

An exhaustive predictive framework was established to predict childhood immunization status and to identify significant associated factors through the application of various machine learning algorithms in Python 3.12.13. The employed models encompassed Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Light Gradient Boosting Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost).

These algorithms were chosen for their complementary strengths in modeling both linear and nonlinear relationships, as well as complex interactions within the diverse pooled DHS dataset. Logistic Regression acted as a traditional baseline, while tree-based, ensemble, distance-based, margin-based, and neural network methodologies offered a range of learning strategies for comparative analysis.

All models underwent training under consistent preprocessing conditions and were assessed using the same train–test split and performance metrics to guarantee an equitable comparison. This multi-model strategy diminished dependence on a single classifier and established a solid foundation for identifying the most effective model.

Model Training

Upon finalizing the data cleaning and preprocessing stages, the analytical dataset comprised 49,817 children aged between 12 and 23 months. This dataset was divided into a training set (80%; n = 39,853) and an independent test set (20%; n = 9,964), with the distribution of childhood immunization status maintained in both subsets. Subsequently, model training was

conducted on the training data utilizing 10-fold cross-validation, a method that entails partitioning the dataset into ten subsets; during each iteration, nine folds were employed for model fitting while one fold was set aside for validation. The performance metrics were averaged across the folds to yield dependable estimates of model performance, robustness, and generalizability.

Model Evaluation

The performance of the model was assessed using a hold-out test dataset. A consistent evaluation framework was implemented across all models, incorporating metrics such as accuracy, precision, recall, specificity, F1-score, ROC-AUC, and PR-AUC. The ROC-AUC metric evaluated the overall discriminative capability, whereas the PR-AUC offered supplementary assessment in the context of class imbalance. Confusion matrices, established at a 0.5 classification threshold, were employed to calculate sensitivity, specificity, positive predictive value, and negative predictive value. To estimate 95% confidence intervals for the performance metrics, bootstrap resampling was conducted on the test dataset with 1,000 iterations. All models adhered to the same preprocessing pipeline, data partitioning strategy, and evaluation criteria to facilitate a fair comparison. The results were interpreted as predictive associations within the aggregated multi-country DHS dataset, rather than as evidence of causation.

Hyper Parameter Tuning

The process of optimizing hyperparameters was undertaken to improve the predictive accuracy and generalizability of the selected machine learning models. The algorithms that were utilized included Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Artificial

Neural Network, Light Gradient Boosting Machine (LightGBM), and eXtreme Gradient Boosting (XGBoost).

The utilization of tree-based and ensemble methods was driven by their capability to effectively capture complex nonlinear relationships and higher-order interactions among predictors, while Logistic Regression was employed as a standard benchmark (74). Furthermore K-Nearest Neighbors, and the Artificial Neural Network were integrated to provide complementary approaches based on margins, distances, and nonlinear learning.

The tuning procedure was conducted exclusively on the training dataset, utilizing randomized search alongside stratified 10-fold cross-validation. The area under the receiver operating characteristic curve (ROC-AUC) was adopted as the primary optimization criterion, due to its dependability in assessing discriminatory performance amidst class imbalance (75). For each model, a specific hyperparameter space was defined, and the configuration that achieved the highest cross-validated ROC-AUC was identified as the final tuned model (76).

Results

A total of 49,817 children aged 12–23 months were included in the analysis. This section describes the distribution of child, maternal, healthcare utilization, and household socioeconomic characteristics of the study population, followed by the analysis of factors associated with childhood immunization status and the predictive performance of the machine learning models.

Child-related characteristics

Among the 49,817 children included in the study, 50.84% were male and 49.16% were female. More than half (53.07%) were aged 12–17 months, while 46.93% were 18–23 months. The

majority of children were born as single births (96.33%), with only 3.67% from multiple births. Regarding birth order, 59.92% were of birth order 1–3, followed by 28.88% with birth order 4–6, and 11.20% with birth order ≥7. Nearly half of the children (48.84%) had a long birth interval (≥48 months), while 38.60% and 12.56% had medium and short birth intervals, respectively. In terms of birth size, 48.89% were reported as average, 31.89% as large, and 19.21% as small. Most children were currently breastfeeding (64.39%). Approximately 28% of children had experienced diarrhea (28.12%), fever (28.07%), or acute respiratory infection (28.06%) in the two weeks preceding the survey (Table 1).

Table 1 . Child-related characteristics

Variable	Categories	Frequency	Percentage
Child's sex	Male	25,327	50.84
	Female	24,490	49.16
Child's age (months)	12–17 months	26,436	53.07
	18–23 months	23,381	46.93
Multiple birth status	Single birth	47,987	96.33
	Multiple birth	1,830	3.67
Birth order	1–3	29,849	59.92
	4–6	14,389	28.88

Variable	Categories	Frequency	Percentage
	≥7	5,579	11.20
Birth interval	Short (<24 months)	6,259	12.56
	Medium (24–47 months)	19,229	38.60
	Long (≥48 months)	24,329	48.84
Birth size	Small	9,571	19.21
	Average	24,358	48.89
	Large	15,888	31.89
Currently breastfeeding status	No	17,740	35.61
	Yes	32,077	64.39
Diarrhea in the last 2 weeks	No	35,809	71.88
	Yes	14,008	28.12
Fever in the last 2 weeks	No	35,834	71.93
	Yes	13,983	28.07
Acute Respiratory Infection (ARI) symptoms	No	35,836	71.94
	Yes	13,981	28.06

Maternal-related characteristics

The majority of mothers (69.72%) fell within the age range of 20 to 34 years, followed by 21.61% who were aged 35 to 49 years, and 8.67% who were between 15 and 19 years. Approximately one-third (33.88%) of these mothers had not received any formal education, while 30.97% had completed primary education, 30.49% had attained secondary education, and a mere 4.66% had pursued higher education. Just over half of the mothers were employed (53.63%), and a significant majority (85.14%) were either married or in a union. More than half (57.52%) experienced their first childbirth before reaching the age of 20. In terms of parity, 42.72% had low parity (1–2 children), 29.14% had moderate parity (3–4 children), and 28.14% had high parity (5 or more children). Furthermore, 51.30% of mothers were literate, whereas 48.70% were not (Table 2).

Table 2 Maternal-related characteristics

Variable	Categories	Frequency	Percentage
Mother's age (years)	15–19 years	4,317	8.67
	20–34 years	34,733	69.72
	35–49 years	10,767	21.61
Mother's education level	No education	16,877	33.88
	Primary	15,429	30.97
	Secondary	15,188	30.49
	Higher	2,323	4.66
Mother's employment status	Not working	23,100	46.37
	Working	26,717	53.63

Mother's marital status	Not currently married/in union	7,403	14.86
	Currently married/in union	42,414	85.14
Age at first birth	<20 years	28,654	57.52
	20–24 years	15,835	31.79
	≥25 years	5,328	10.70
Parity (number of children ever born)	Low parity (1–2 children)	21,282	42.72
	Moderate parity (3–4 children)	14,518	29.14
	High parity (≥5 children)	14,017	28.14
Mother's literacy status	Not literate	24,261	48.70
	Literate	25,556	51.30

Maternal healthcare utilization characteristics

In terms of maternal healthcare utilization, 64.75% of mothers participated in four or more antenatal care (ANC) visits, while 25.58% attended 1–3 visits, and 9.67% did not attend any ANC visits. A significant majority of births were overseen by skilled health personnel (76.59%) and took place in health facilities (74.68%). Conversely, 23.41% of births occurred without skilled attendance, and 25.32% took place at home. Only one-third of mothers (33.05%) received postnatal care, leaving 66.95% without such care. A considerable number of mothers (78.87%)

lacked health insurance coverage. Furthermore, 35.48% indicated that the distance to a health facility posed a significant challenge, whereas 64.52% did not view it as a major obstacle (Table 3).

Table 3 Maternal healthcare utilization characteristics

Variable	Categories	Frequency	Percentage
Antenatal care (ANC) visits	No ANC visit	4,817	9.67
	1–3 visits	12,744	25.58
	≥4 visits	32,256	64.75
Skilled birth attendance	No	11,661	23.41
	Yes	38,156	76.59
Place of delivery	Home	12,613	25.32
	Health facility	37,204	74.68
Postnatal care received	No	33,354	66.95
	Yes	16,463	33.05
Health insurance coverage	No	39,289	78.87
	Yes	10,528	21.13
Distance to health facility	Not a big problem	32,144	64.52

Variable	Categories	Frequency	Percentage
	Big problem	17,673	35.48

Household socioeconomic, environmental, and information characteristics

Regarding In terms of household socioeconomic, environmental, and informational characteristics, 27.29% of households were categorized within the lowest wealth quintile, followed by 21.64% in the lower category, 19.99% in the middle category, 17.13% in the upper category, and 13.97% in the highest category. A considerable proportion of participants, 66.91%, resided in rural areas, while 33.09% lived in urban settings. Household sizes were generally large, with 37.83% of households consisting of 5–7 members, 34.64% having 8 or more members, and 27.53% comprising 1–4 members. The majority of households were led by males, representing 77.84%. In terms of religious affiliation, 51.41% identified as Protestant or belonging to other Christian denominations, 33.79% as Catholic, 10.65% as traditional or other/no religion, and 4.15% as Muslim. Regarding the educational attainment of partners, 36.65% had achieved secondary education or higher, 35.59% had completed primary education, and 27.76% had no formal education.

Access to essential services revealed that 65.42% of households had access to improved drinking water, while 46.70% had access to enhanced sanitation facilities. Electricity was available to 38.83% of households. A significant majority, 85.37%, depended on polluting cooking fuels, whereas only 14.63% used clean fuels. In terms of information access, 59.54% of mothers had exposure to mass media, while 40.46% did not. Internet usage was particularly low, with only 24.99% of individuals reporting internet access compared to 75.01% who did not (Table 4).

555 **Table 4 Household socioeconomic, environmental, and information characteristics**

Variable	Categories	Frequency	Percentage
Wealth index	poorest	13,593	27.29
	poorer	10,778	21.64
	Middle	9,956	19.99
	richer	8,532	17.13
	richest	6,958	13.97
Place of residence	Urban	16,486	33.09
	Rural	33,331	66.91
Household size	1–4	13,715	27.53
	5–7	18,846	37.83
	8+	17,256	34.64
Sex of household head	Male	38,778	77.84
	Female	11,039	22.16
Religion	Catholic	16,833	33.79
	Protestant/Other Christian	25,612	51.41

Variable	Categories	Frequency	Percentage
	Islam	2,065	4.15
	Traditional/Other/No religion	5,307	10.65
Partner's education level	No education	13,831	27.76
	Primary education	17,728	35.59
	Secondary or higher	18,258	36.65
Source of Drinking water	Unimproved	17,227	34.58
	Improved	32,590	65.42
Toilet facility	Unimproved	26,551	53.30
	Improved	23,266	46.70
Electricity access	No	30,473	61.17
	Yes	19,344	38.83
Cooking fuel type	Polluting fuel	42,531	85.37
	Clean fuel	7,286	14.63
Mass media exposure	No	20,157	40.46

Variable	Categories	Frequency	Percentage
	Yes	29,660	59.54
Internet use	No	37,368	75.01
	Yes	12,449	24.99
Literacy Status	NO	24,261	48.70
	Yes	25,556	51.30

Immunization Coverage Among Children Aged 12–23 Months in 22 Sub-Saharan African Countries

The pooled analysis encompassing 49,817 children aged 12 to 23 months across 22 Sub-Saharan African nations revealed significant disparities in vaccination coverage for individual vaccine doses. In total, 40.1% of children were fully immunized, while 59.9% were not, underscoring a notable shortfall in achieving comprehensive childhood immunization coverage in the region (Figure 2).

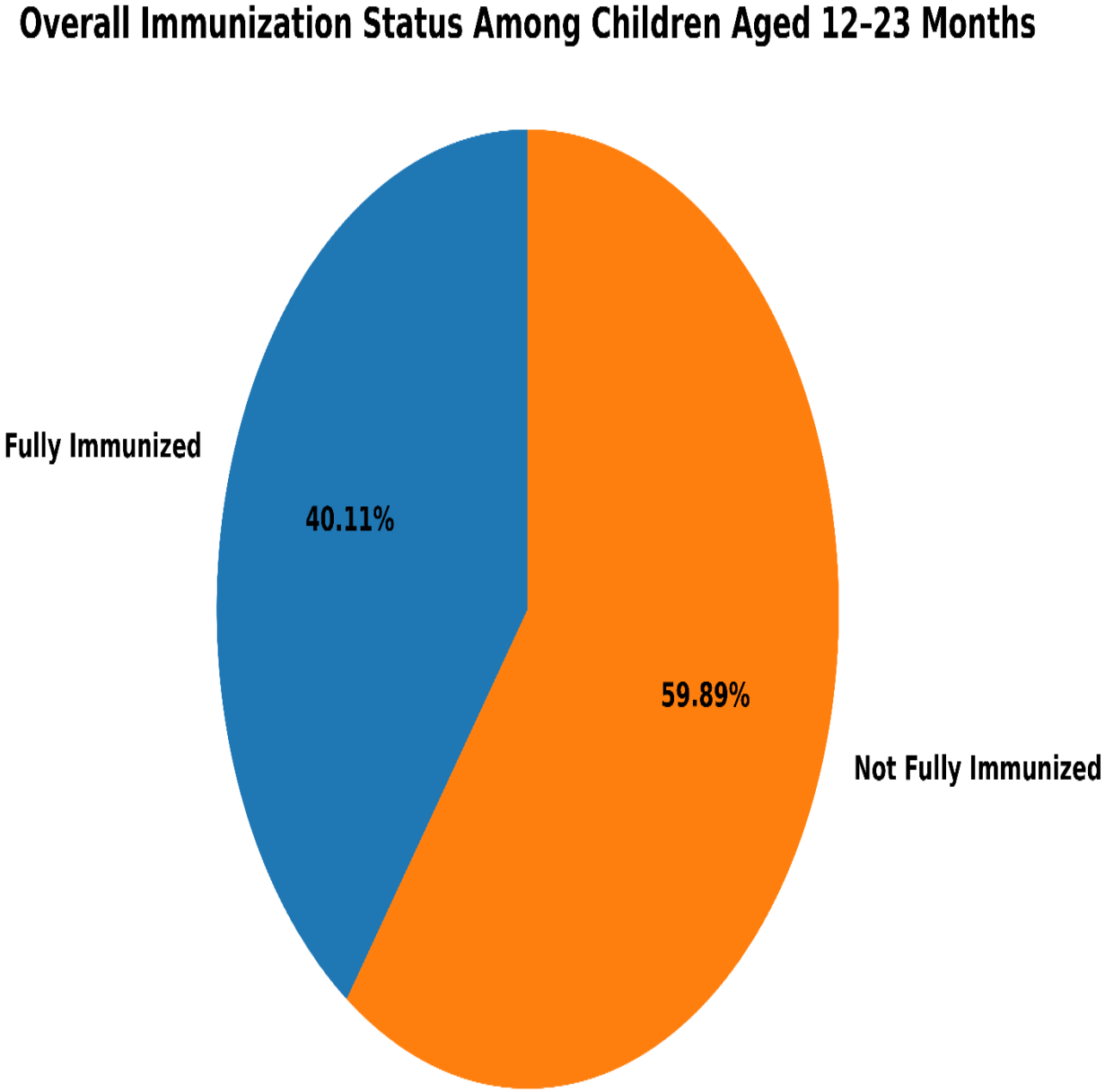
Among the various vaccine antigens, BCG vaccination exhibited the highest coverage at 80.6%, followed by Polio-1 at 78.7%, Pentavalent-1 at 77.9%, and PCV-1 at 74.9%. These results indicate a relatively high acceptance of vaccines administered during early immunization visits,

particularly for first-dose vaccines. However, there was a progressive decline in coverage with subsequent doses, revealing a significant dropout rate along the vaccination schedule.

The most pronounced decline was noted in the pentavalent vaccine series, which fell from 77.9% for Pentavalent-1 to 59.9% for Pentavalent-3, reflecting an 18 percentage-point dropout. Likewise, polio vaccination rates decreased from 78.7% for Polio-1 to 67.8% for Polio-3, and PCV coverage dropped from 74.9% for PCV-1 to 64.4% for PCV-3. Measles vaccination coverage stood at 67.7%, indicating that nearly one-third of children were left unprotected against measles.

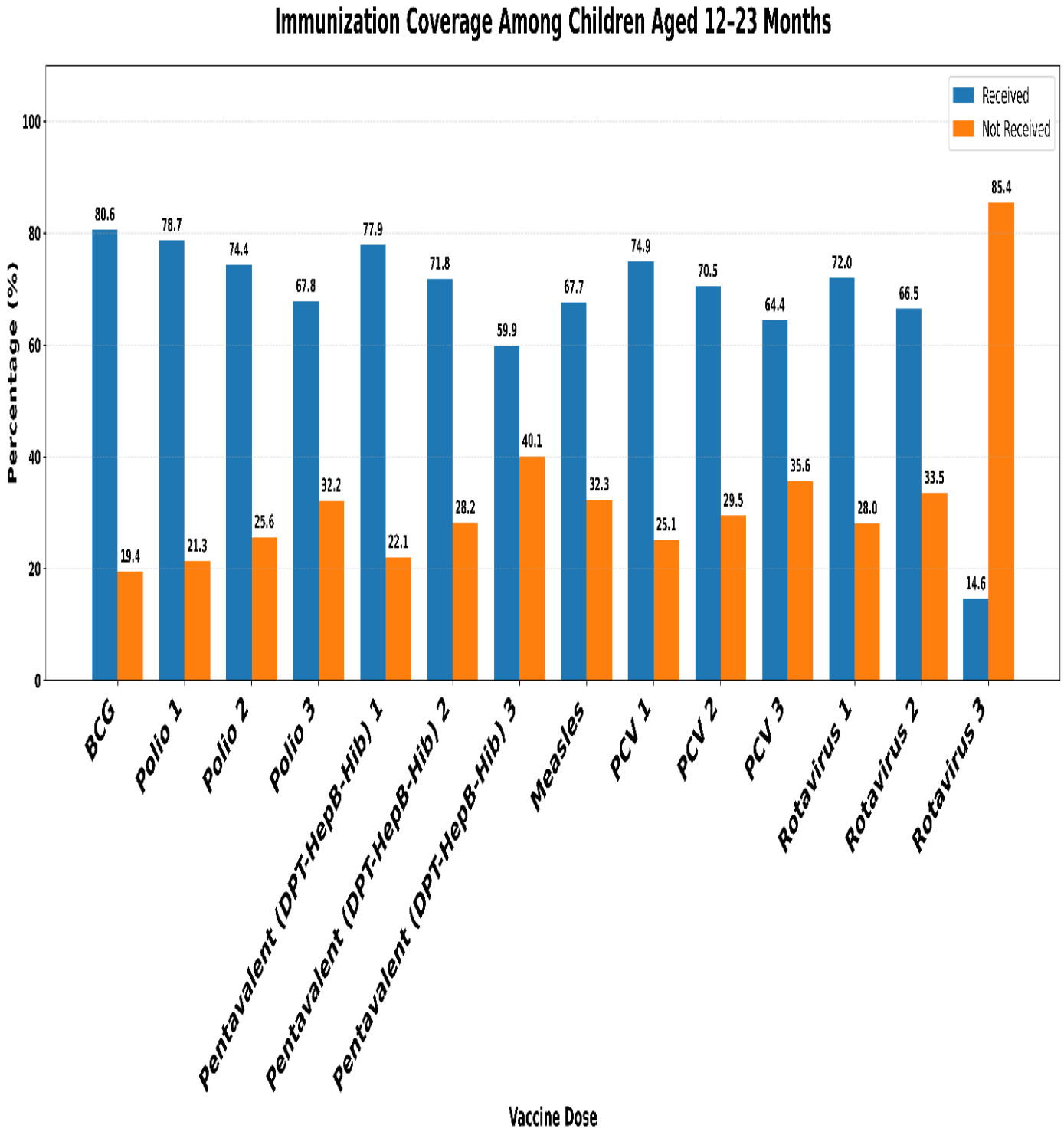
The most significant immunization gap was identified for the rotavirus vaccine, particularly concerning the third dose. While 72.0% of children received Rotavirus-1, coverage decreased to 66.5% for Rotavirus-2 and plummeted to 14.6% for Rotavirus-3. This drastic reduction may reflect issues related to vaccine availability, adherence to multi-dose schedules, access to health systems, and caregiver follow-up (Figure 3).

In summary, the findings illustrate that although initial engagement with immunization services is relatively robust in Sub-Saharan Africa, ensuring that children complete the entire vaccination schedule poses a considerable challenge. The high rates of partially immunized and unimmunized children emphasize the urgent need for interventions aimed at promoting timely vaccinations.



587

588 **Figure 2 Prevalence of childhood immunization**



589

590 **Figure 3. Coverage of individual vaccine antigens among children aged 12–23 months in 22**
591 **Sub-Saharan African countries**
592

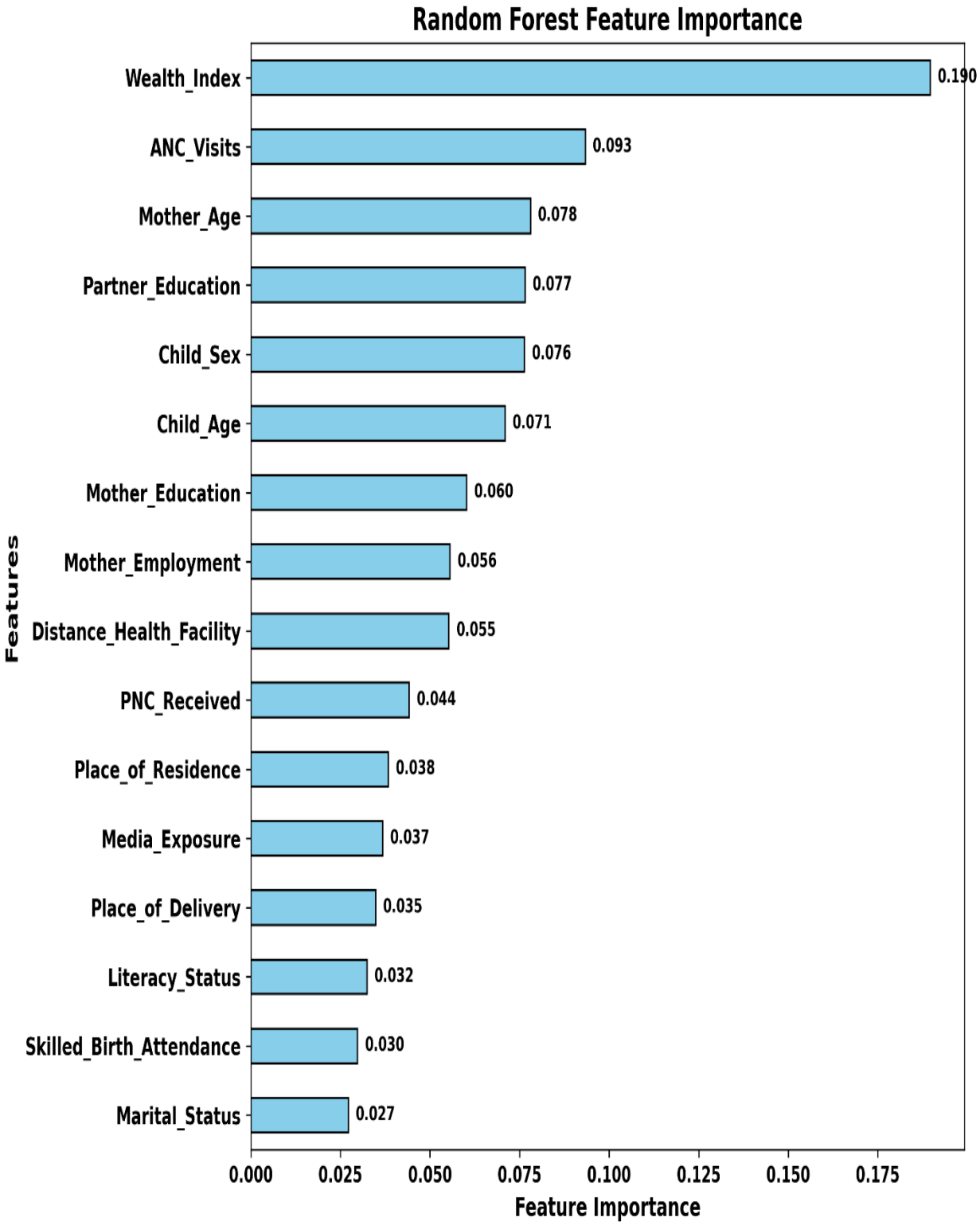
593 **Importance feature selection**

Feature selection is an essential component of machine learning, as it enhances parsimony, reduces redundancy, and improves both predictive performance and interpretability (77, 78). In this investigation, 44 candidate variables were evaluated, and Recursive Feature Elimination (RFE) was employed to retain the most informative predictors by iteratively discarding variables that contributed the least to the model's performance (79).

The RFE process was validated using tenfold cross-validation, with accuracy, precision, recall, F1-score, and AUC calculated at each iteration to ensure that feature removal did not compromise predictive performance.

Subsequent this, Random Forest was utilized to rank the importance of variables based on each predictor's contribution to model accuracy (80). The most significant predictors included household wealth index, ANC visit, mother age, partner education, child sex, child age, mother education and mother employment (Figure 4). These variables were retained for subsequent modeling and interpretation.

The synergistic use of RFE and RF enabled a robust, data-driven feature selection process, resulting in a streamlined and high-performing predictive model for childhood immunization status (81). Since these methods are data-driven, the selected predictors should be interpreted as significant for prediction rather than as causal factors.



611

612 **Figure 4 Important features selected by recursive feature elimination (RFE)**

613

Class Balancing

The analysis of childhood immunization status revealed a significant class imbalance, with one immunization category markedly outnumbering the other within the pooled dataset. While this observation mirrors the actual epidemiological landscape of childhood immunization, it poses a risk of biasing model training towards the majority class, thereby hindering the identification of children belonging to the minority class. To counteract this issue, the Synthetic Minority Oversampling Technique (SMOTE) was exclusively applied to the training dataset (82). SMOTE generates synthetic minority-class observations through interpolation between existing cases, thus improving class representation without simple duplication and supporting the development of more robust predictive model learning (83, 84).

The implementation of SMOTE occurred after one-hot encoding of categorical variables, including country, with k=5 employed to decrease the likelihood of generating implausible synthetic observations. The balanced training set that resulted improved the model's ability to discern patterns within the minority class. Sensitivity analyses performed on the original imbalanced dataset revealed comparable performance trends and feature importance rankings, indicating that SMOTE improved minority class detection without significantly distorting the underlying data structure. However, because the data were pooled across multiple countries, synthetic resampling should be interpreted cautiously, as it may not fully preserve country-specific heterogeneity. Consequently, the balanced data were used solely for model training, and the findings are interpreted as predictive patterns rather than population-level estimates.

Development and Performance Comparisons of Machine Learning Models

Multiple supervised machine learning algorithms were systematically developed and evaluated to predict childhood immunization status. To improve model generalizability and optimize predictive performance, a robust machine learning pipeline incorporating tenfold cross-validation and RandomizedSearchCV-based hyperparameter tuning was implemented. The evaluated algorithms included Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), XGBoost, LightGBM (LGBM), K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN). All models were trained using the SMOTE-balanced training dataset and subsequently evaluated using the original imbalanced test dataset to ensure unbiased estimation of real-world predictive performance.

Model performance was assessed using accuracy, precision, recall (sensitivity), F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). To quantify uncertainty and evaluate the stability of model estimates, 95% confidence intervals (CIs) were calculated using bootstrap resampling with 1,000 iterations. Given the imbalanced nature of childhood immunization outcomes, threshold-independent metrics, including ROC-AUC and precision–recall area under the curve (PR-AUC), were considered important indicators for comparative model evaluation.

Among all evaluated models, Random Forest demonstrated the superior predictive performance, achieving the highest accuracy (94.12%; 95% CI: 92.80–95.30), precision (95.05%; 95% CI: 93.60–96.20), recall (92.21%; 95% CI: 90.10–94.00), and F1-score (92.57%; 95% CI: 90.80–94.10). Decision Tree also demonstrated strong predictive ability, with an accuracy of 92.36%

(95% CI: 90.50–93.90), precision of 91.07%, recall of 90.01%, and F1-score of 90.54% (95% CI: 88.20–92.60).

XGBoost achieved moderate predictive performance, with an accuracy of 87.34% (95% CI: 85.00–89.40), precision of 88.95%, recall of 84.35% (95% CI: 81.70–86.80), and F1-score of 70.61% (95% CI: 67.00–74.20). The relatively lower F1-score compared with precision and recall suggests reduced balance between correctly identifying immunized and non-immunized children. K-Nearest Neighbors and LightGBM demonstrated acceptable performance, achieving accuracies of 84.81% (95% CI: 82.10–87.10) and 82.21% (95% CI: 79.40–84.80), respectively.

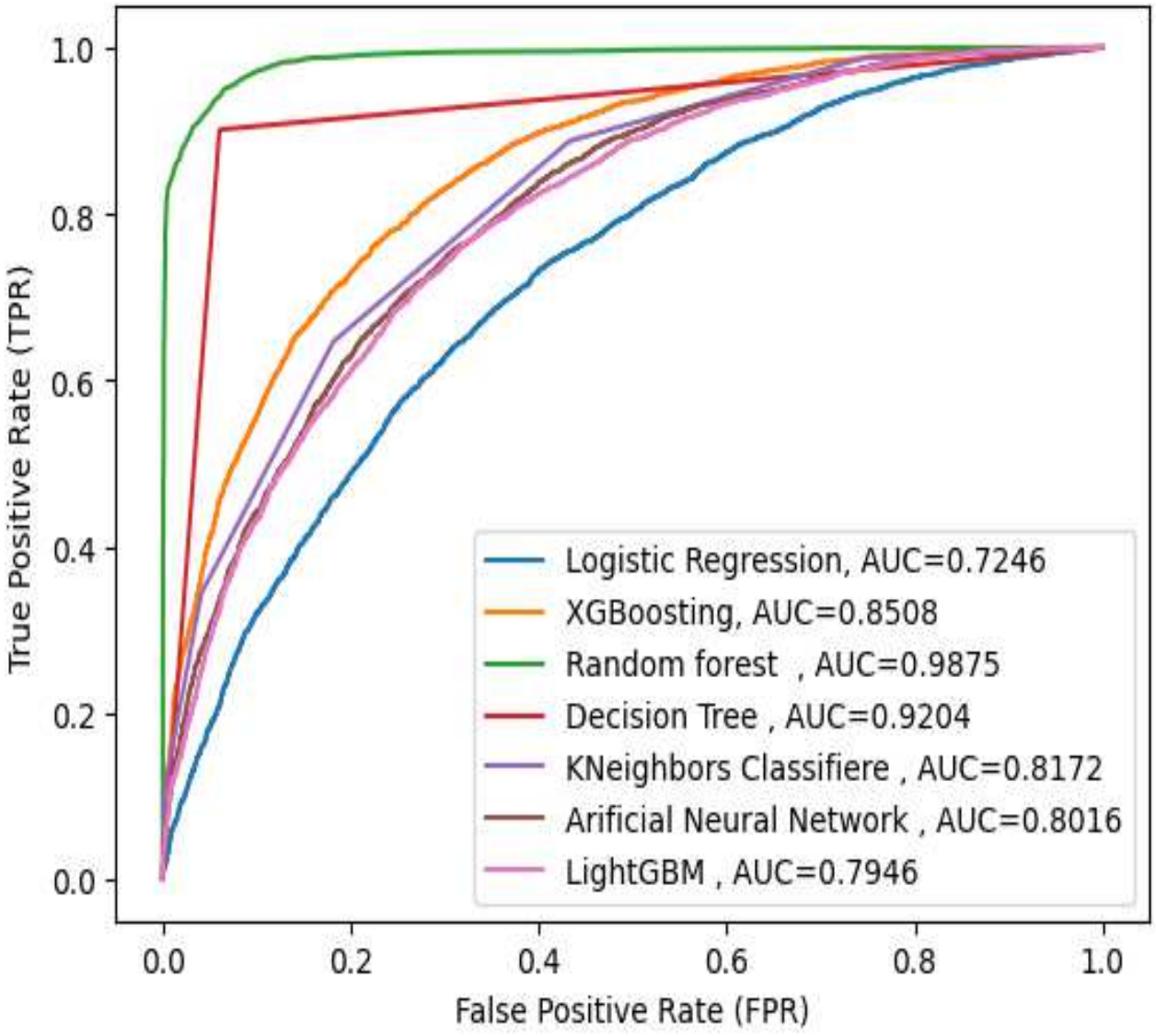
Artificial Neural Network showed comparatively lower predictive performance, with accuracies of 77.46% (95% CI: 74.20–80.30) and 79.65% (95% CI: 76.50–82.40) and Logistic Regression exhibited the lowest overall performance, with an accuracy of 77.29% (95% CI: 74.00–80.10), suggesting limited capability in capturing complex nonlinear interactions among determinants of childhood immunization. The comparative performance results of the developed machine learning models are summarized in Table 5. The discriminative performance of the models is further demonstrated through receiver operating characteristic (ROC) curves (Figure 5) and precision–recall (PR) curves (Figure 6).

Overall, Random Forest consistently outperformed all other models across discrimination, classification, and stability metrics. The narrow confidence intervals observed for Random Forest indicate high precision and robustness of its predictive estimates. This superior performance is likely attributable to the ensemble architecture of Random Forest, which combines multiple decision trees to reduce variance, minimize overfitting, and effectively capture nonlinear relationships and complex interactions among predictors. Although Decision

Tree achieved competitive performance, its single-tree structure may be more susceptible to variance and overfitting. Similarly, boosting-based approaches such as XGBoost and LightGBM demonstrated improved ability to model complex patterns compared with traditional statistical approaches. In contrast, Logistic Regression showed comparatively lower performance due to its assumption of linear relationships between predictors and immunization outcomes. These findings, supported by ROC-AUC, PR-AUC, sensitivity, and balanced accuracy assessments, demonstrate the effectiveness of ensemble machine learning approaches, particularly Random Forest, for predicting childhood immunization status in multi-country DHS data.

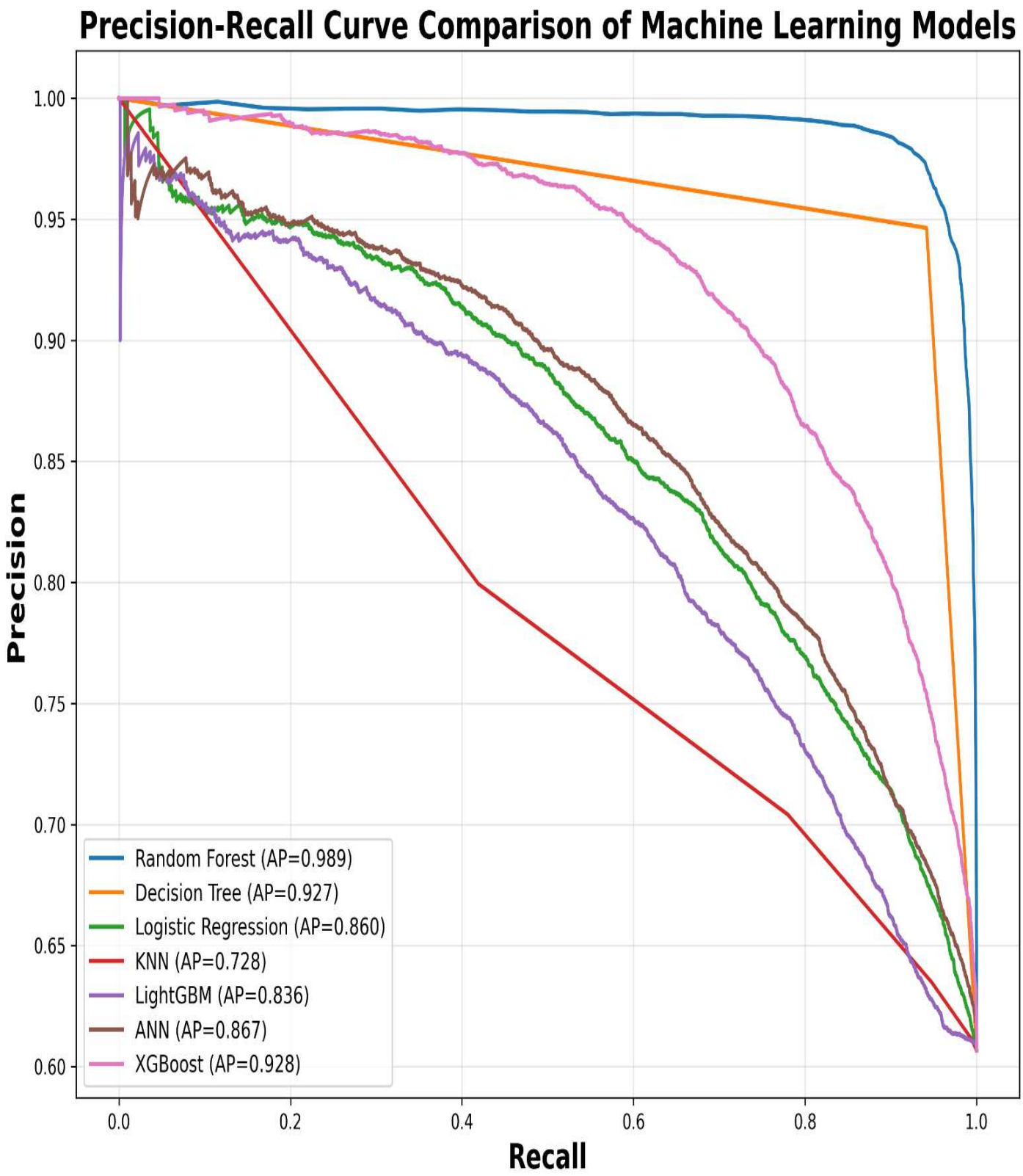
Table 5: Final Model Performance Comparison Table for the machine learning algorithm

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Random Forest (RF)	94.12%	95.05%	92.21%	92.57%
Decision Tree (DT)	92.36%	91.07%	90.01%	90.54%
Logistic Regression (LR)	77.29%	70.12%	80.43%	81.55%
XGBoost	87.34%	88.95%	84.35%	70.61%
K-Nearest Neighbors (KNN)	84.81%	79.23%	83.46%	81.29%
Artificial Neural Network	79.65%	75.77%	86.51%	72.41%
LightGBM (LGBM)	82.21%	78.23%	83.17%	78.45%



692

693 **Figure 5 ROC-AUC curve score of machine learning model**



694

695 **Figure 6 Precision–Recall (PR) curve comparison of machine learning models**

The calibration plot effectively illustrates the alignment between the predicted probabilities and the observed childhood immunization outcomes. The calibration curves of the evaluated machine learning models were compared with the 45° perfect calibration line to assess their agreement between predicted and observed probabilities. Among the evaluated models, the Random Forest (RF) model showed a calibration curve that closely followed the perfect calibration line across most probability ranges, indicating good calibration performance. Minor deviations at some intervals suggest slight overestimation or underestimation of risk; however, these differences are minimal. The low Brier score of the RF model (0.048) further confirms good calibration performance, demonstrating that the model provides reliable and well-calibrated probability estimates for predicting childhood immunization status, as presented in Figure 7.

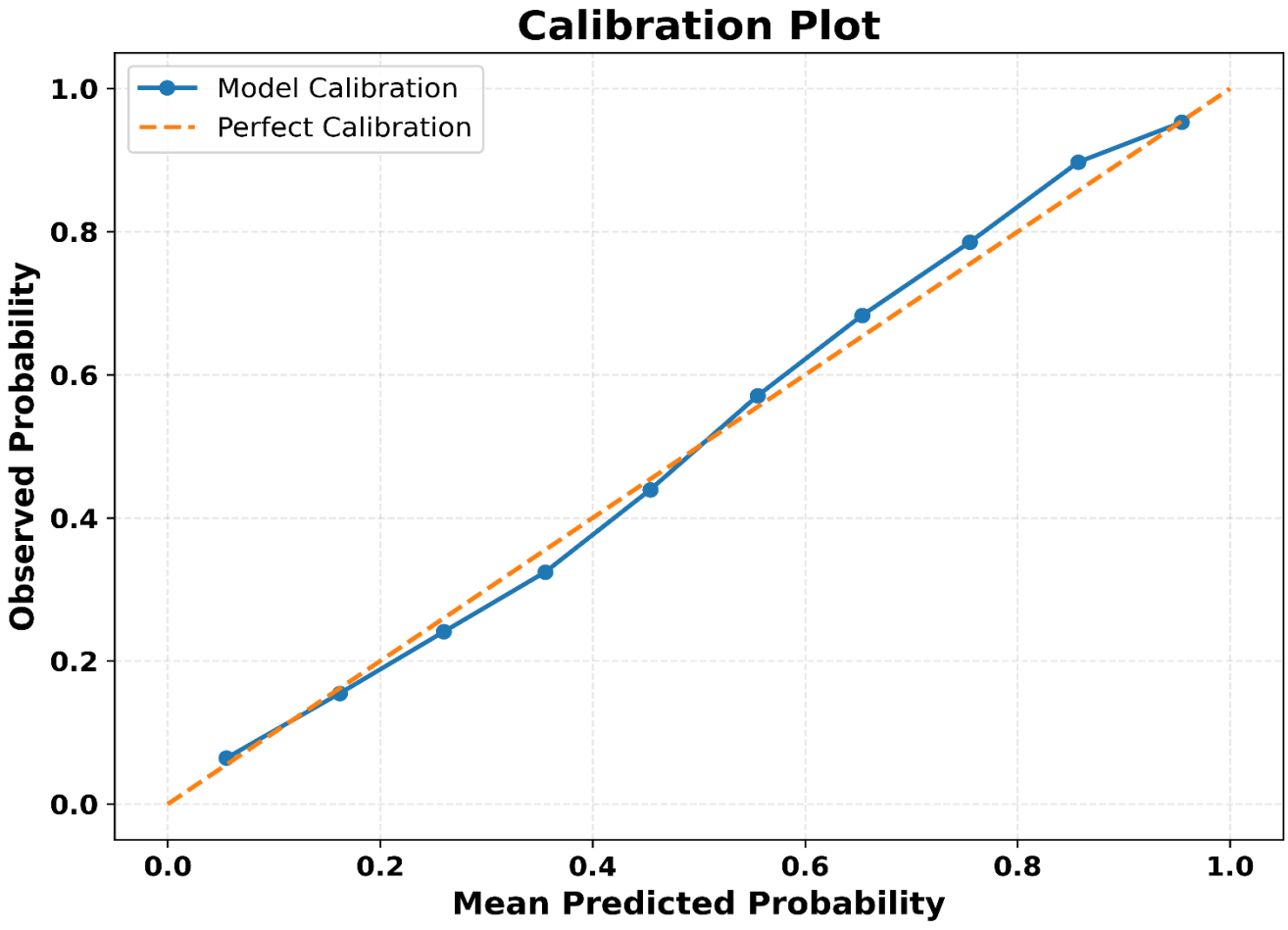


Figure 7 Model calibration plot illustrating the agreement between predicted probabilities and observed childhood immunization outcomes.

SHAP Feature Impact on Model Prediction

As a post hoc interpretability method, SHAP was employed to clarify the predictions produced by the trained Random Forest model, offering a comprehensive framework for decomposing individual predictions into additive contributions from each feature, thus enhancing the transparency of model behavior (85). SHAP values were calculated on the independent test dataset to preserve unbiased interpretation of the fitted model.

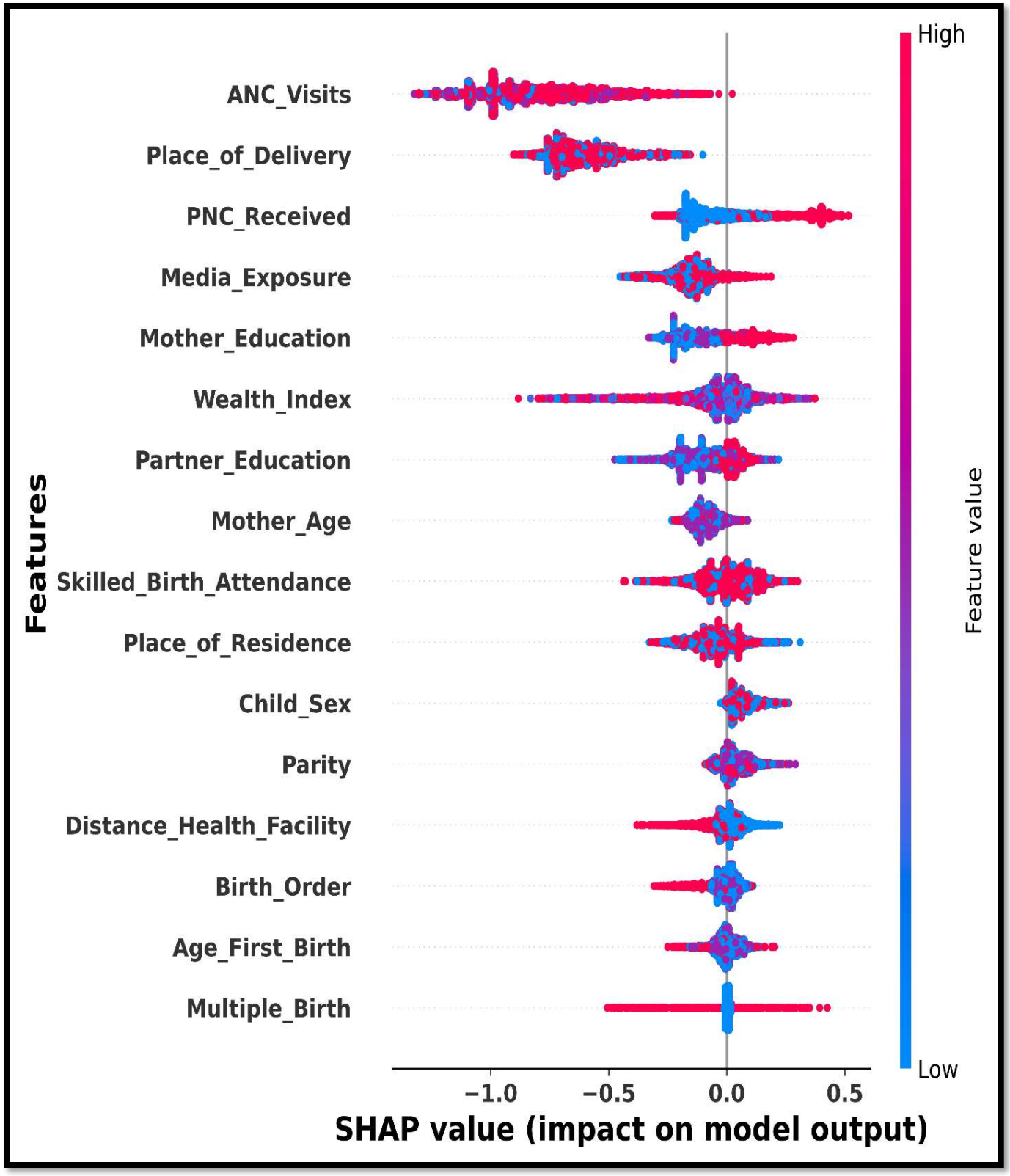
The SHAP analysis revealed that a combination of maternal, household, healthcare utilization, and community-level characteristics contributed to predicting childhood immunization status. Healthcare-related factors showed prominent contributions, with antenatal care (ANC) attendance, skilled birth attendance, and institutional delivery among the most influential predictors of model output. Maternal and socioeconomic characteristics, including maternal education, partner education, household wealth index, exposure to media, maternal age, and place of residence, also showed substantial contributions to the predicted probability of complete immunization. These findings indicate that the Random Forest model captured complex relationships between access to healthcare services, maternal characteristics, and childhood vaccination outcomes (figure 8).

Positive SHAP values denote features that enhanced the model's predicted probability of a child being classified as fully immunized, while negative SHAP values signify features that shifted predictions towards a not fully immunized status. However, this directionality reflects the internal logic of the model rather than a causal or clinical relationship (86). The SHAP summary

plot further revealed that the effect of a given predictor was not uniform across individuals, suggesting nonlinear patterns and feature interactions learned by the model.

To enhance interpretability, SHAP-based findings were complemented with partial dependence plots and feature interaction analyses. These additional outputs helped explore marginal effects and interaction patterns without altering the underlying data structure. Given that SHAP values are model-specific and may be affected by correlated predictors and the pooled multi-country design, they should be viewed as relative predictive contributions rather than causal effects.

Overall, SHAP improved the transparency of the Random Forest model by clarifying how different features contributed to prediction. This interpretability framework helps bridge the gap between predictive accuracy and explainability, which is particularly valuable for evidence-informed public health decision-making in multi-country settings.



741

742 **Figure 8 SHAP feature impact on model prediction**

Model Interpretation (Higher Predicted Probability of Full Immunization)

The SHAP waterfall plot illustrates how individual socioeconomic, demographic, and healthcare-related factors contribute to a specific model prediction regarding full childhood immunization. The model's baseline prediction, $E[f(X)] = 0.424$, was elevated to the final prediction, $f(x) = 5.027$, as a result of the cumulative positive impacts of the selected features.

Among the evaluated features, the richest household wealth category exhibited the most significant positive impact on the prediction of full immunization. Other important positive contributors included access to electricity, higher maternal education, internet use, clean cooking fuel, urban residence, maternal literacy, and maternal employment status. These positive SHAP values indicate that these features enhanced the model's prediction, increasing the likelihood of full immunization for the individual child (figure 9).

Overall, this waterfall plot highlights that improved socioeconomic conditions, access to information, and maternal empowerment-related factors were important contributors to the model’s prediction of childhood immunization status. However, these findings should be interpreted as explanations of model behavior rather than causal relationships; SHAP values indicate how each feature influenced the prediction generated by the model, not whether the factor directly causes immunization completion.

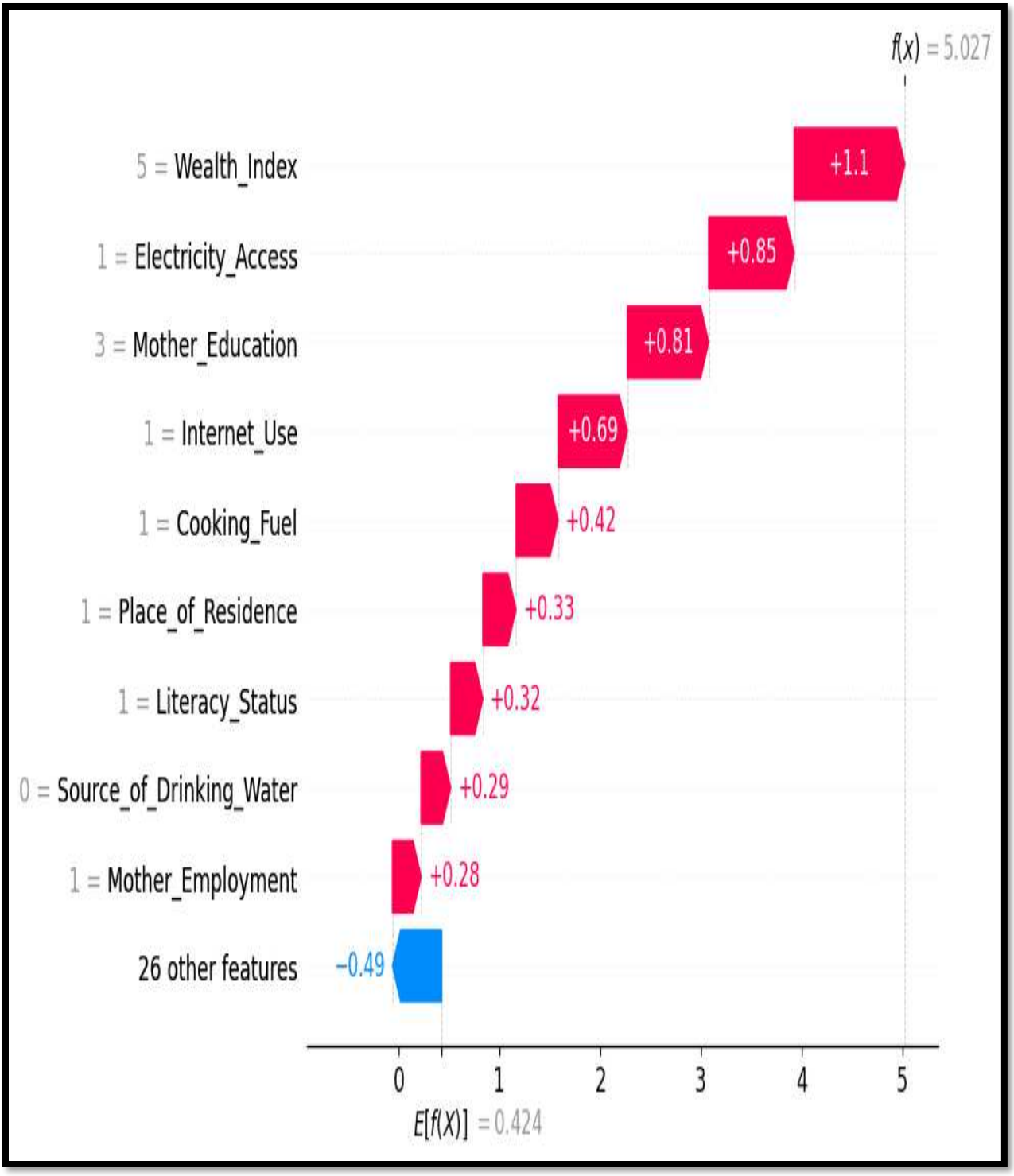


Figure 9 SHAP waterfall plot for higher predicted probability of full childhood immunization

Model Interpretation (Lower Predicted Probability of Full Immunization)

The SHAP waterfall plot illustrates the extent to which individual features influenced the model's prediction, transitioning the output from the baseline *(E[f(X)] = 0.424)* to a diminished final prediction (f(x) = −0.982), which signifies a decreased probability of complete childhood immunization.

This reduced prediction is mostly driven by several adverse socioeconomic and maternal factors that negatively impacted the model's output. The lack of access to electricity had the most significant detrimental effect, followed by factors such as maternal unemployment, low educational attainment of partners, maternal illiteracy, absence of maternal education, and the lowest household wealth status. Collectively, these elements shifted the prediction away from full immunization, highlighting the cumulative detriments linked to limited resources, inadequate educational levels, and restricted access to supportive environments. While certain factors—such as possessing health insurance and receiving postnatal care (PNC)—had a positive influence on the prediction, their impact was relatively minor and insufficient to offset the more substantial negative contributions (figure 10).

This indicates that standalone healthcare interventions may fall short in the context of broader socioeconomic challenges, underscoring the necessity for integrated and multisectoral strategies.

In summary, the SHAP analysis reveals that socioeconomic deprivation and factors related to limited maternal empowerment were pivotal in diminishing the model's predicted probability of complete childhood immunization in this instance, whereas healthcare access factors offered only limited alleviation.

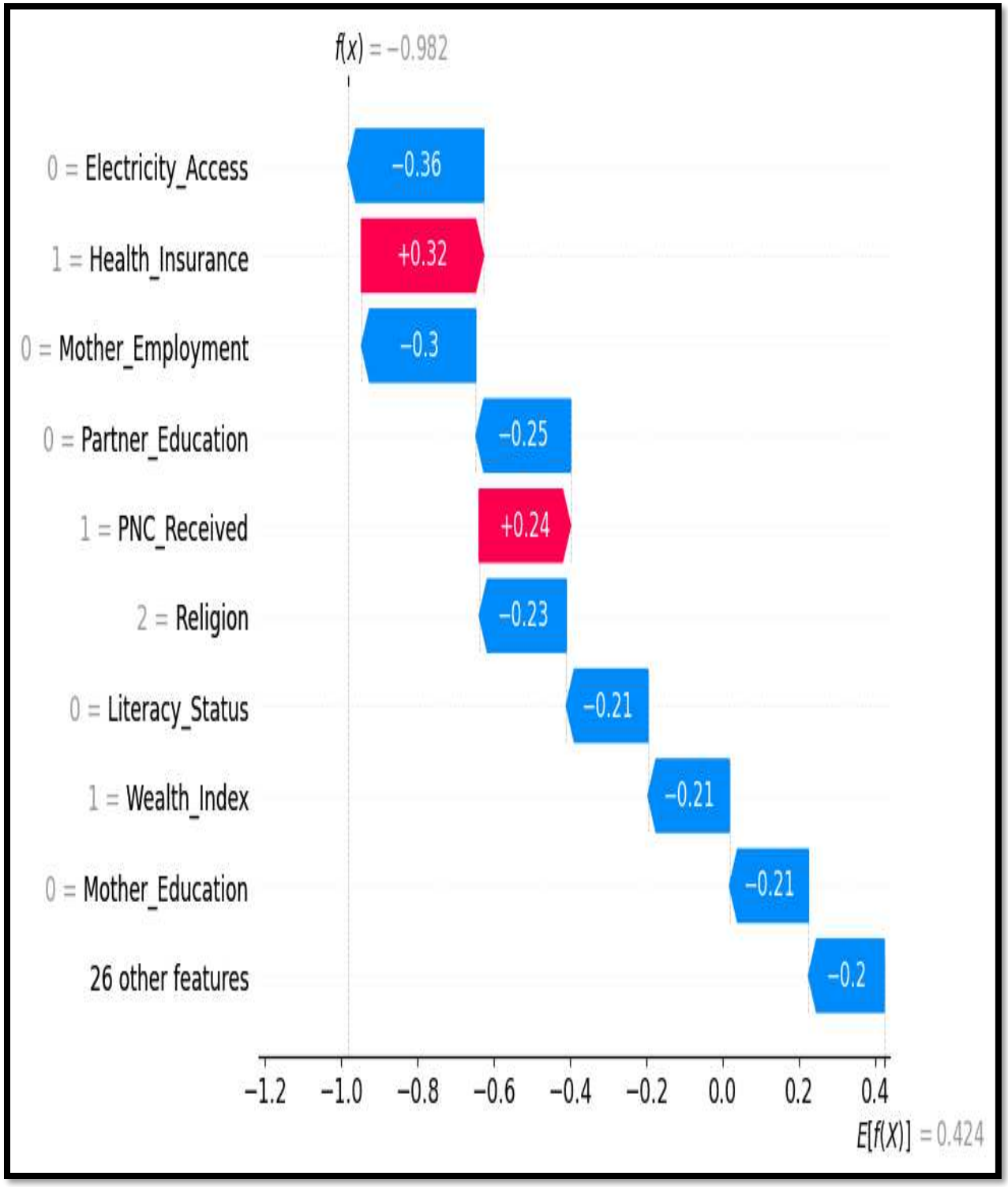


Figure 10 SHAP waterfall plot for lower predicted probability of full childhood immunization

Discussion

This study provides a comprehensive analysis of childhood immunization status and its associated factors across Sub-Saharan Africa (SSA), leveraging recent Demographic and Health Survey (DHS) data from 2019 to 2025 across 22 countries. A pooled sample of children aged 12–23 months was analyzed, incorporating child-related, maternal-related, maternal healthcare utilization, household socioeconomic, environmental, and information characteristics as predictors. Seven supervised machine learning algorithms—Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), XGBoost, LightGBM (LGBM), and Artificial Neural Network (ANN)—were developed and rigorously evaluated using 10-fold cross-validation. Recursive Feature Elimination (RFE) was employed to identify the most informative predictors before model development, and model interpretation approaches were applied to quantify the contribution of key determinants to childhood immunization prediction.

Among the developed models, the Random Forest (RF) algorithm demonstrated the highest predictive performance, achieving an accuracy of 94.12%, followed by other evaluated machine learning models, including Decision Tree (DT), XGBoost, K-Nearest Neighbors (KNN), LightGBM (LGBM), Artificial Neural Network (ANN), and Logistic Regression (LR). The superior performance of the RF model highlights the ability of ensemble-based approaches to effectively capture complex, nonlinear relationships and interactions among influencing childhood immunization status.

These findings are consistent with previous studies (87, 88) demonstrating the effectiveness of Random Forest algorithms in predicting health outcomes using population-based survey data, particularly when multiple interacting determinants contribute to the outcome.

The superior performance of RF may be attributed to its ensemble learning mechanism, which combines multiple decision trees to improve generalizability, reduce overfitting, and capture complex patterns within heterogeneous population-level health data. This capability is particularly valuable for childhood immunization prediction, where vaccination status is influenced by diverse factors.

To improve model transparency and interpretability, Explainable Artificial Intelligence (XAI) techniques were incorporated. At the global level, feature importance analysis was used to quantify the relative contribution of predictors and identify the major determinants associated with childhood immunization status across the study population. At the individual level, SHapley Additive exPlanations (SHAP) analysis illustrated the direction and magnitude of each predictor's contribution to specific model predictions, providing insight into why individual children were classified as fully or not fully immunized. This combined global and local interpretation framework enhances the interpretability, reliability, and practical utility of machine learning models for informing targeted immunization interventions and evidence-based public health decision-making.

It is important to note that feature importance and SHAP-based explanations represent model-derived predictive contributions and should not be interpreted as causal relationships or direct measures of epidemiological importance. These interpretations may also be influenced by correlations among predictors and the heterogeneous nature of pooled multi-country DHS data.

The analysis identified several important predictors of childhood immunization status across Sub-Saharan Africa. Key contributing factors included maternal healthcare utilization characteristics, particularly antenatal care (ANC) attendance, place of delivery, and postnatal

834 care (PNC) utilization, which reflecting the importance of continued engagement with health
835 services. Maternal-related characteristics, including education and media exposure, also
836 contributed substantially to model predictions, highlighting the role of maternal knowledge and
837 access to health information. In addition, household socioeconomic characteristics, particularly
838 wealth status, place of residence, and partner education, were important predictors, while child-
839 related and geographic characteristics further contributed to the prediction of immunization
840 outcomes. These findings underscore the multifactorial nature of childhood immunization status
841 across diverse settings in Sub-Saharan Africa.

842 Antenatal care (ANC) utilization emerged as the most influential predictor of childhood
843 immunization status in this study, and this finding is consistent with previous studies in Benin
844 (37, 38),Ghana (39-41), Indonesian (42),Ethiopia (43), Tanzania (44) ,East Arica (45-48) ,and
845 sub-Saharan Africa (20, 49), which similarly reported a positive association between ANC
846 attendance and childhood immunization coverage, with a clear gradient effect showing that
847 mothers who attended more ANC visits(four and above ANC visits) were more likely to have
848 fully immunized children, suggesting a dose–response relationship between ANC utilization and
849 immunization uptake (89).

850 Antenatal care (ANC) provides repeated and structured contact between pregnant women and the
851 health system, creating important opportunities for health education, early counseling, and the
852 development of trust between women and providers (90, 91). During these visits, mothers can be
853 informed about the importance of childhood vaccination, the recommended immunization
854 schedule, and the need for timely completion of all vaccine doses, thereby increasing the
855 likelihood of continued engagement with postnatal and child health services and ultimately
856 improving immunization uptake (92, 93).

Moreover, ANC attendance is often linked with other positive health-seeking behaviors, including institutional delivery and postnatal care (PNC) utilization, which further reinforce adherence to vaccination schedules (94, 95). Women who attend antenatal care (ANC) are more likely to receive tetanus toxoid vaccination during pregnancy, reflecting prior acceptance of vaccines and greater confidence in health interventions, and thus ANC serves not only as a source of information but also as a marker of broader healthcare access and utilization (96, 97).

The strong association between ANC and childhood immunization likely reflects underlying socioeconomic and structural advantages, as women who attend ANC are more likely to be educated, live closer to health facilities, and have better access to health information, thereby reducing barriers such as low awareness, distance, and missed vaccination opportunities (89, 98). Programmatically, these findings highlight the importance of using ANC as a strategic platform for immunization counseling, reminder systems, and follow-up mechanisms, particularly in low-resource and disadvantaged settings (99).

Place of delivery was a significant factor in childhood immunization status, with children born in health facilities more likely to be fully immunized than those delivered at home, and this pattern is consistent with prior studies from Ethiopia (100-104) ,Kenya (105), Benin (38) , Ugandan (106), East Africa (47),Afghanistan (107), India (108), and Sub-Saharan Africa (20, 49). This may be explained by the immediate linkage of facility-born children to essential newborn services, including early vaccination, counseling on immunization schedules, and better follow-up through health systems (109, 110). In contrast, home deliveries often limit early contact with skilled health providers, reducing opportunities for timely vaccination and health education (111). Furthermore, facility delivery reflects broader maternal healthcare utilization and health-seeking behavior, as women who deliver in health facilities are more likely to access skilled care,

postnatal services, and continued engagement with child health programs, ultimately improving immunization uptake (112, 113). In many low-resource settings, women who deliver in health facilities are also more likely to continue using postnatal and child health services, reflecting broader engagement with the continuum of care (114, 115). These differences suggest that place of delivery is not only a birth outcome but also a marker of maternal healthcare access and health-seeking behavior, as facility delivery is often associated with greater use of skilled care, postnatal services, and child health programs that contribute to higher immunization uptake (116).

Household wealth index emerged as a significant determinant of childhood immunization in this study, with children from wealthier households more likely to be fully immunized than those from poorer households. This finding is consistent with evidence from Ethiopia (28, 117-119), Benin (38) ,Indonesian (42) and across sub-Saharan Africa (20, 49), where higher socioeconomic status is associated with better access to health services and more complete vaccination coverage. The relationship may reflect the fact that wealthier households are better able to overcome financial and geographic barriers, afford transportation, and access health information through media and other sources (120, 121). In addition, they are more likely to seek care promptly, attend health facilities, and absorb both direct and indirect vaccination-related costs such as transport, lost time, and service expenses (122). These findings suggest that economic disadvantage remains an important barrier to equitable immunization coverage and highlight the need for targeted interventions for poorer households (123).

Place of residence was identified as a significant predictor of childhood immunization status, underscoring persistent geographic disparities in access to vaccination services. This observation aligns with earlier studies conducted in Nigeria (124, 125), Ethiopia (126-128), Ghana (129),

East Africa (47), and sub-Saharan Africa (20, 130), which consistently reported lower immunization completion among rural children, largely due to constraints in healthcare access, service availability, and delivery systems.

Children living in rural areas may have lower immunization coverage due to longer distances to health facilities, limited transportation, fewer outreach services, and reduced access to health information (131, 132). In contrast, urban children often benefit from better healthcare infrastructure, greater availability of vaccination services, and more frequent contact with healthcare providers (133, 134). These findings emphasize the need for targeted immunization strategies, including strengthening rural outreach programs, expanding community-based vaccination services, and improving equitable access to routine immunization (135).

Postnatal care (PNC) utilization was an important determinant of childhood immunization status in this study, with children whose mothers received PNC more likely to be fully immunized than those whose mothers did not, consistent with evidence from Sudan (136), Ethiopia (117, 137), Benin (38) ,West African countries (138), East Africa (46, 47),Nigeria (139), and Sub-Saharan Africa (20, 48). This association may be explained by the continued contact between mothers and health providers after delivery, which provides opportunities for immunization counseling, vaccination reminders, and reinforcement of recommended child health practices (140). PNC visits also facilitate early identification of newborn health problems and strengthen linkage to routine immunization services (141, 142).

Furthermore, PNC utilization reflects broader maternal healthcare engagement and health-seeking behavior, as mothers who attend postnatal services are more likely to use other preventive child health interventions, including immunization programs (143, 144). In contrast,

lack of PNC may result in missed opportunities for health education, delayed vaccination, and reduced continuity of care (145). These findings emphasize the importance of integrating immunization promotion and follow-up mechanisms into postnatal care platforms, particularly in settings where access to routine child health services remains limited (146).

Media exposure emerged as an important determinant of childhood immunization status in this study, consistent with previous studies from Ethiopia (28, 147), Ghana (41), East Africa (47, 148), Indonesia (149), and Sub-Saharan Africa (20, 49), demonstrating that children whose mothers had greater exposure to mass media were more likely to be fully immunized than those whose mothers had limited or no media exposure, possibly due to improved access to vaccination information, increased awareness of immunization schedules, and enhanced utilization of basic child health services (150, 151). This association may be explained by the role of media in increasing awareness of vaccination schedules, correcting misconceptions, and reminding caregivers about the importance of timely child health services (152, 153). In settings where access to health information is limited, media platforms such as radio, television, and mobile messaging can serve as effective channels for promoting immunization uptake and encouraging appropriate health-seeking behavior (154, 155).

Media exposure can also reinforce positive health behaviors by increasing knowledge, addressing vaccine misconceptions, and encouraging timely use of preventive health services (156). In many low-resource settings, where access to health information remains unequal, mass media serves as an important channel for reaching caregivers and promoting immunization practices (99). However, limited media exposure may contribute to missed vaccination opportunities due to inadequate awareness and delayed healthcare-seeking behavior (153). These findings suggest that strengthening targeted health communication strategies through accessible

media platforms could support improvements in childhood immunization coverage, particularly among disadvantaged populations (157, 158).

Maternal education emerged as a significant determinant of childhood immunization, with children of more educated mothers being more likely to be fully vaccinated than those whose mothers had limited or no formal education, consistent with evidence from Ethiopia (117, 126, 159), Eritrea (160), Ghana (129), Somalia (161), Nigeria (162), East Africa (163), and across Sub-Saharan Africa (20), where higher maternal educational attainment is strongly associated with improved vaccination uptake. This relationship may reflect the role of education in enhancing maternal health literacy, awareness of vaccination schedules, healthcare-seeking behavior, and decision-making capacity for child health (164, 165). Educated mothers are more likely to understand the benefits of immunization, recognize the importance of timely vaccination, communicate effectively with health providers, and access and utilize preventive health services (166, 167). Conversely, limited maternal education may reduce awareness of immunization benefits and increase the risk of missed vaccination opportunities due to inadequate knowledge or challenges in navigating health services (168, 169). These findings highlight the importance of promoting female education and implementing tailored health communication strategies to improve immunization coverage among children of mothers with lower educational attainment (170).

Partner education was positively associated with childhood immunization in this study, indicating that higher paternal or partner educational attainment may contribute to better vaccination coverage among children Likewise, fathers who have attended primary education and higher were more likely to fully vaccinate their children than fathers who did not attend formal education. It was supported by studies reported in Pakistan (171), Somalia (172), West

971 African countries (138) and sub-Saharan Africa (20),This association may reflect the role of
972 educated partners in strengthening household decision-making, increasing awareness of the
973 importance and schedule of immunization, and promoting timely health-seeking behavior (173).
974 In many settings, fathers or partners influence whether caregivers attend health facilities, comply
975 with vaccination appointments, and prioritize preventive child health services (174). Educated
976 partners may also be better positioned to understand health messages, support the use of child
977 health services, and reduce barriers to immunization through improved communication and
978 informed support within the household (175).

979 Maternal age was significantly associated with childhood immunization status, with older
980 mothers more likely to have fully immunized children than younger mothers. T Similar findings
981 have been reported in Ethiopia (176-178), Tanzania (179), Nigeria (180), Myanmar (181),
982 Senegal (182), East Africa (148) and A scoping review (183) ,which reported that increasing
983 maternal age is associated with improved childhood immunization coverage. This association
984 may be explained by the greater experience, confidence, and accumulated knowledge of older
985 mothers regarding child healthcare practices, vaccination schedules, and the importance of
986 preventive health services (167). Older mothers may also have greater autonomy in household
987 decision-making and improved ability to navigate health systems, which can facilitate timely
988 vaccination uptake (184). In contrast, younger mothers may face challenges such as limited
989 health knowledge, reduced decision-making power, lower exposure to health information, and
990 fewer experiences with routine child health services, contributing to missed or delayed
991 vaccinations (183). Strengthening targeted immunization education and support for younger
992 mothers, including during antenatal and postnatal care contacts, may help reduce age-related
993 disparities in vaccination uptake (185).

Skilled birth attendance emerged as a significant predictor of childhood immunization status, with higher odds of full immunization among children born to mothers who received assistance from skilled birth attendants. This association may be explained by the increased interaction between mothers and healthcare providers during skilled deliveries, which provides opportunities for health education, vaccination counseling, registration of newborns, and linkage to routine immunization services (186). In addition, skilled birth attendants can facilitate early postnatal care engagement and reinforce adherence to age-appropriate vaccination schedules (37). Consistent associations have been documented in Guinea (187), Benin (37), Indonesian (42), East Africa (148), highlighting the role of skilled delivery in improving access to essential child health services, including immunization. Expanding skilled delivery coverage and embedding immunization promotion within maternal and newborn care could help increase childhood vaccine uptake (188).

Overall, this study demonstrates the value of machine learning approaches in predicting childhood immunization status and identifying key determinants across Sub-Saharan Africa. The strong predictive performance of the Random Forest model highlights the ability of data-driven methods to capture complex interactions among maternal, child, household, healthcare, information, and geographic factors that influence vaccination completion. The findings emphasize that improving childhood immunization coverage requires strengthening maternal healthcare platforms, particularly ANC, delivery care, and PNC services, while addressing socioeconomic inequalities, improving health information access, and promoting equitable utilization of preventive health services. By integrating predictive modeling with explainable artificial intelligence techniques, this study provides actionable insights that can support targeted interventions, prioritize high-risk groups, and inform evidence-based immunization strategies in

resource-limited settings. Future research incorporating longitudinal data, additional health system indicators, and country-specific analyses may further enhance the generalizability and practical application of machine learning models for immunization programs.

Conclusion and Recommendations

This study demonstrates that machine learning approaches can effectively predict childhood immunization status among children aged 12–23 months using pooled DHS data from 22 Sub-Saharan African countries. Among the evaluated models, the Random Forest algorithm achieved the highest predictive performance, demonstrating its ability to capture complex and nonlinear relationships among child-related, maternal-related, maternal healthcare utilization, household socioeconomic, environmental, and information characteristics influencing immunization status. Explainable Artificial Intelligence (XAI) analysis identified adequate antenatal care (ANC) utilization (four or more ANC visits) as the most influential predictor of childhood immunization status. Other important determinants included institutional delivery, receipt of postnatal care (PNC), skilled birth attendance, higher maternal education, higher partner education, older maternal age, urban residence, higher household wealth status, and media exposure, which contributed positively to the prediction of full immunization among children.

These findings highlight that childhood immunization status is shaped by interconnected individual, household, healthcare, and contextual factors rather than a single determinant. The integration of machine learning with explainable approaches provides valuable insights into immunization patterns and key factors, thereby supporting evidence-based decision-making to improve vaccination coverage across Sub-Saharan Africa.

Based on these findings, immunization programs should strengthen the integration of vaccination promotion within maternal healthcare platforms, particularly ANC, delivery care, and PNC services, to maximize opportunities for counseling, follow-up, and timely vaccination. Targeted interventions should prioritize mothers and families with limited education, low media exposure, and socioeconomic disadvantages through tailored health communication strategies, community-based outreach, and improved access to reliable immunization information. Efforts should also focus on reducing geographic and financial barriers by expanding equitable access to quality immunization services, especially in underserved communities. Furthermore, the application of machine learning-based prediction models, supported by explainable AI techniques, can assist health programs and policymakers in identifying children with suboptimal immunization status, optimizing resource allocation, and designing targeted immunization interventions.

Future research should incorporate longitudinal data, additional health system indicators, and country-specific analyses to further improve model generalizability and support context-specific strategies for achieving equitable childhood immunization coverage in Sub-Saharan Africa.

Strengths and Limitations

This study has several important strengths. First, it utilized a large pooled DHS dataset from 22 Sub-Saharan African countries, providing a comprehensive regional assessment of childhood immunization status and enhancing the regional applicability of the findings across diverse settings. Second, the study incorporated a wide range of predictors, including child-related, maternal-related, maternal healthcare utilization, household socioeconomic and environmental, information-related, and geographic characteristics, allowing a multidimensional assessment of factors influencing immunization completion. Third, the application of seven machine learning

algorithms, including Random Forest, Decision Tree, K-Nearest Neighbor, Logistic Regression, Artificial Neural Network, XGBoost, and LightGBM, enabled a robust comparison of predictive performance and identification of the most suitable modeling approach. Fourth, the use of Recursive Feature Elimination (RFE), cross-validation, and Explainable Artificial Intelligence (XAI) techniques such as SHAP analysis improved model reliability, reduced redundant predictors, and enhanced interpretation of key predictors. Finally, the study provides actionable insights by identifying high-impact predictors that can support targeted immunization interventions and evidence-based public health decision-making.

Despite these strengths, the study has some limitations. First, the cross-sectional nature of DHS data limits the ability to establish temporal relationships and causal inferences between predictors and childhood immunization status. Second, vaccination information was obtained from vaccination cards and maternal recall; therefore, recall bias and misclassification of immunization status may be possible, particularly in settings where vaccination records are incomplete. Third, although the pooled multi-country approach improves statistical power and regional applicability, differences in healthcare systems, immunization policies, socioeconomic conditions, and vaccination practices across countries may introduce heterogeneity that could influence model performance. Fourth, machine learning models were developed primarily for prediction rather than causal explanation; therefore, feature importance and SHAP values should be interpreted as model-derived predictive contributions rather than direct evidence of epidemiological causation. Fifth, the analysis did not include some potentially relevant factors, such as vaccine supply availability, quality of health facility services, caregiver attitudes, vaccine confidence, and community-level health system indicators, due to limitations in DHS data availability. Future studies incorporating longitudinal designs, routine immunization data, and

additional health system variables are recommended to further improve prediction accuracy and support context-specific immunization strategies.

List of abbreviations

ANC: Antenatal Care;

ANN: Artificial Neural Network;

BCG: Bacillus Calmette–Guérin;

CI: Confidence Interval;

CV: Cross-Validation;

DHS: Demographic and Health Survey;

DT: Decision Tree;

F1-score: F1-score (Harmonic Mean of Precision and Recall);

IA2030: Immunization Agenda 2030;

KNN: K-Nearest Neighbors;

LGBM: Light Gradient Boosting Machine;

LR: Logistic Regression;

ML: Machine Learning;

PCV: Pneumococcal Conjugate Vaccine;

PNC: Postnatal Care;

PR-AUC: Precision–Recall Area Under the Curve;

RF: Random Forest;

RFE: Recursive Feature Elimination;

ROC-AUC: Receiver Operating Characteristic–Area Under the Curve;

SHAP: SHapley Additive exPlanations;

SMOTE: Synthetic Minority Over-sampling Technique;

SSA: Sub-Saharan Africa;

WHO: World Health Organization;

XAI: Explainable Artificial Intelligence.

Declarations

Ethical approval and consent to participate

The study was based on publicly available data from the Demographic and Health Surveys (DHS) Program. Permission to access and use the datasets was obtained from the DHS Program website (http://www.dhsprogram.com) after submitting a brief study description. Ethical approval for DHS data collection was obtained by the DHS Program from the ICF Institutional Review Board (IRB) and from the national ethics review committees of each participating country. The surveys were conducted in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants by the DHS Program prior to data collection. The datasets were anonymized, and no personal identifiers were used in this secondary analysis. No additional ethical approval was required for this study.

Data Availability Statement

The data used in this study are publicly available from the Demographic and Health Surveys (DHS) Program. Access to the datasets can be obtained by registered users through the DHS Program website (http://www.dhsprogram.com) upon reasonable request and approval of a brief study description.

Funding

Not applicable

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Consent for publication

Not applicable

Competing Interests

All authors declared that they have no any competing interest.

Clinical trial number: not applicable**.**

Authors' contribution

MFT was responsible for conceptualization, design, data extraction, formal analysis, language editing, data interpretation, methodology, funding acquisition, project administration, resource management, software development, supervision, data curation, article review, validation, visualization, and writing the original draft as well as reviewing and editing.

SFT,DKA,DEA,MEZ,AEA,BAM,HNA,KYW,TAA,ABG,JMK,HAK,AKM,AEG,TCM,and FBK contributed to data curation, conceptualization, formal analysis, and funding acquisition. They reviewed the study design and draft manuscript, verified the analysis, and made significant contributions through critical revision and article review. All authors approved the final version of the manuscript

Acknowledgments

We gratefully acknowledge the DHS Program for granting access to the DHS datasets used in this study. The data were utilized in accordance with the DHS research ethics and data use guidelines.

References

1. Zhou F, Jatlaoui TC, Leidner AJ, Carter RJ, Dong X, Santoli JM, et al. Health and economic benefits of routine childhood immunizations in the era of the vaccines for children program—United States, 1994–2023. MMWR Morbidity and Mortality Weekly Report. 2024;73.
2. Lindstrand A, Cherian T, Chang-Blanc D, Feikin D, O'Brien KL. The world of immunization: achievements, challenges, and strategic vision for the next decade. The Journal of infectious diseases. 2021;224(Supplement_4):S452-S67.
3. Nigatu AM, Aschalew Z, Estifanos W, Kasse T, Gemede A, Getie A. Timely vaccination coverage and associated factors among children aged 12–23 months in Arbaminch Town, Ethiopia, 2024. BMC Public Health. 2026;26(1):1973.
4. Melaku T, Gudina EK, Sørensen JB, Dræbel TA, Meyrowitsch DW, Kavishe RA, et al. Trends and recovery of routine childhood immunization before, during, and after the COVID-19 pandemic in Ethiopia: A five-year trend analysis. Vaccine. 2026;77:128343.
5. Shattock AJ, Johnson HC, Sim SY, Carter A, Lambach P, Hutubessy RCW, et al. Contribution of vaccination to improved survival and health: modelling 50 years of the Expanded Programme on Immunization. The Lancet. 2024;403(10441):2307-16.
6. Bura VK, Srivastava R, Chawla AK, Njambe TOE, Khanal S, Sangal L, et al. Essential programme on immunization in WHO South-East Asia: A five-decade journey of saving millions of lives & ending diseases. The Indian Journal of Medical Research. 2024;160(3-4):267.
7. Thomas N. Vaccines and immunization. World Health Organization (WHO). 2024.
8. Ali HA, Hartner A-M, Echeverria-Londono S, Roth J, Li X, Abbas K, et al. Vaccine equity in low and middle income countries: a systematic review and meta-analysis. International journal for equity in health. 2022;21(1):82.
9. Dimitrova A, Carrasco-Escobar G, Richardson R, Benmarhnia T. Essential childhood immunization in 43 low-and middle-income countries: Analysis of spatial trends and socioeconomic inequalities in vaccine coverage. PLoS medicine. 2023;20(1):e1004166.
10. Papelian S. Global and national trends in childhood vaccination coverage from 1980 to 2023 with forecasts to 2030. 2025.

11. Oyugi B, Kallander K, Shahabuddin A. Strengthening Primary Health Care Through Implementation Research: Strategies for Reaching Zero-Dose Children in Low-and Middle-Income Countries' Immunization Programs. Vaccines. 2025;13(10):1040.
12. Jethwaney J, Dhalaria P. Immunization Ecosystem: Perspectives and Strategies: Taylor & Francis; 2026.
13. Golla EB, Geremew H, Abate A, Ali MA, Simegn MB, Tilahun WM, et al. Vaccination dropout and associated factors among children in Ethiopia: a systematic review and meta-analysis (2014-2024). BMC pediatrics. 2025;25(1):426.
14. Lubanga AF, Bwanali AN, Kangoma M, Matola Y, Moyo C, Kaonga B, et al. Addressing the re-emergence and resurgence of vaccine-preventable diseases in Africa: A health equity perspective. Human Vaccines & Immunotherapeutics. 2024;20(1):2375081.
15. Sinumvayo JP, Munezero PC, Tope AT, Adeyemo RO, Bale MI, Mutsaka-Makuvaza MJ, et al. Vaccination and vaccine-preventable diseases in Africa. Scientific African. 2024;24:e02199.
16. Jean Simon D, Kondo Tokpovi VC, Dianou K, Okonji OC, Kiragu A, Olorunsaiye CZ, et al. Regional, subregional and country-level full vaccination coverage in children aged 12–23 months for 34 countries in sub-Saharan Africa: a global analysis using demographic and health survey data. BMJ Global Health. 2025;10(3):e018333.
17. Okesanya OJ, Olatunji G, Olaleke NO, Mercy MO, Ilesanmi AO, Kayode HH, et al. Advancing immunization in Africa: overcoming challenges to achieve the 2030 global immunization targets. Adolescent Health, Medicine and Therapeutics. 2024:83-91.
18. Ivanova V, Shahabuddin A, Sharkey A, Johri M. Advancing immunization coverage and equity: A structured synthesis of pro-equity strategies in 61 Gavi-supported countries. Vaccines. 2023;11(1):191.
19. Hussein SA, Osman MM, Hassan MM, Hassan YSA, Hussein AA, Mohamed AH, et al. The future impact of zero-dose children in inaccessible conflict-affected areas of Somalia: aligned with the immunization agenda 2030. Trop Med Health. 2025;53(1):162.
20. Fenta SM, Biresaw HB, Fentaw KD, Gebremichael SG. Determinants of full childhood immunization among children aged 12–23 months in sub-Saharan Africa: a multilevel analysis using Demographic and Health Survey Data. Tropical Medicine and Health. 2021;49(1):29.
21. Bangura JB, Xiao S, Qiu D, Ouyang F, Chen L. Barriers to childhood immunization in sub-Saharan Africa: A systematic review. BMC public health. 2020;20(1):1108.
22. Mano O, Musuka G, Cuadros DF. Immunization coverage, equity, and access for children with disabilities: a scoping review of challenges, strategies, and lessons learned to reduce the number of zero-dose children. 2025.
23. Alemayehu B. Vaccination Coverage and Associated factors among Children in Southwest Ethiopia Peoples' Region. Journal of Health and Biomedical Science. 2026;1(1):21-30.
24. Rodrigues CM, Plotkin SA. Impact of vaccines; health, economic and social perspectives. Frontiers in microbiology. 2020;11:1526.
25. Ma S, Lavelle TA, Ollendorf DA, Lin PJ. Herd Immunity Effects in Cost-Effectiveness Analyses among Low- and Middle-Income Countries. Applied health economics and health policy. 2022;20(3):395-404.
26. Decouttere C, De Boeck K, Vandaele N. Advancing sustainable development goals through immunization: a literature review. Globalization and Health. 2021;17(1):95.

27. Tamir TT, Agimas MC, Angaw DA. Prevalence, spatial patterns and determinants of zero-dose vaccination among children aged 12–23 months in Sub-Saharan Africa: A multilevel spatial analysis. PLOS ONE. 2026;21(3):e0344293.
28. Fenta SM, Fenta HM. Individual and community-level determinants of childhood vaccination in Ethiopia. Archives of public health = Archives belges de sante publique. 2021;79(1):53.
29. Sarder MA, Lee KY, Keramat SA, Hashmi R, Ahammed B. A multilevel analysis of individual and community-level factors associated with childhood immunisation in Bangladesh: Evidence from a pooled cross-sectional survey. Vaccine: X. 2023;14:100285.
30. Yeboah D, Engmann G, Jakperik D. Modelling the determinants of uptake of complete immunization of children in Ghana. BMC pediatrics. 2025;25.
31. Carrodano C. General framework of nonlinear factor interactions using bayesian networks for risk analysis applied to road safety and public health. Scientific Reports. 2025;15(1):29934.
32. Hajiyev A, Xu J. High-Dimensional Regression Analysis and Artificial Intelligence. Springer Books. 2026.
33. Latha N, Geetha K, Damodharan S. Overview of Regression Models and How to Determine the Best Model for Data. Journal of Scientific Research and Reports. 2024;30:250-66.
34. Zhao Q, Zhang C, Zhang W, Zhang S, Liu Q, Guo Y. Applications and challenges of biomarker-based predictive models in proactive health management. Frontiers in public health. 2025;13:1633487.
35. Shu X, Ye Y. Knowledge Discovery: Methods from data mining and machine learning. Social Science Research. 2023;110:102817.
36. Rane NL, Paramesha M, Choudhary SP, Rane J. Machine learning and deep learning for big data analytics: A review of methods and applications. Partners Universal International Innovation Journal. 2024;2(3):172-97.
37. Budu E, Seidu A-A, Agbaglo E, Armah-Ansah EK, Dickson KS, Hormenu T, et al. Maternal healthcare utilization and full immunization coverage among 12–23 months children in Benin: a cross sectional study using population-based data. Archives of Public Health. 2021;79(1):34.
38. Budu E, Ahinkorah BO, Guets W, Ameyaw EK, Essuman MA. Socioeconomic and residence-based related inequality in childhood vaccination in Sub-Saharan Africa: Evidence from Benin. 2023;6(4):e1198.
39. Sem S, Sarfo JO, Amoadu M, Hormenu T. Uptake and determinants of routine vaccines among children aged 12–23 months in adansi South district of Ghana. a cross-sectional study. BMC pediatrics. 2025;25(1):198.
40. Kumar AD, Gidisu JA. Logistic Regression Modeling for Determinants of Childhood Vaccination Uptake in Ghana: A Population-Based Study. International Journal of Interdisciplinary Research in Arts and Humanities. 2025;10(1):86-97.
41. Tekeba B, Tamir TT, Zegeye AF. Prevalence and determinants of full vaccination coverage according to the national schedule among children aged 12–35 months in Ghana. Scientific Reports. 2025;15(1):13.
42. Efendi F, Pradiptasiwi DR, Krisnana I, Kusumaningrum T, Kurniati A, Sampurna MTA, et al. Factors associated with complete immunizations coverage among Indonesian children aged 12–23 months. Children and Youth Services Review. 2020;108:104651.

43. Nour TY, Farah AM, Ali OM, Osman MO, Aden MA, Abate KH. Predictors of immunization coverage among 12-23 month old children in Ethiopia: systematic review and meta-analysis. BMC Public Health. 2020;20(1):1803.
44. Shakiru TH, Mukyanuzi EN, Kinyota DS. Bayesian spatial analysis of child vaccination coverage and associated determinants in Tanzania. BMJ Public Health. 2026;4(2):e004130.
45. Endalamew SG, Gebeye E, Assefa SK, Geremew BM. Immunization Coverage and Its Determinants Among Children Aged 12–23 Months in East Africa: A Bayesian Hierarchical Modeling Approach Based on DHS 2019–2022. BioMed Research International. 2025;2025(1):6645541.
46. Dires AA, Workie DL, Teklie AK. Exploring factors influencing childhood immunization status in East Africa using multilevel ordinal logistic regression analysis. Frontiers in Public Health. 2025;Volume 12 - 2024.
47. Tesema GA, Tessema ZT, Tamirat KS, Teshale AB. Complete basic childhood vaccination and associated factors among children aged 12–23 months in East Africa: a multilevel analysis of recent demographic and health surveys. BMC Public Health. 2020;20(1):1837.
48. Budu E, Ahinkorah BO, Aboagye RG, Armah-Ansah EK, Seidu A-A, Adu C, et al. Maternal healthcare utilsation and complete childhood vaccination in sub-Saharan Africa: a cross-sectional study of 29 nationally representative surveys. BMJ Open. 2021;11(5):e045992.
49. Dzinamarira T, Mano O, Musuka G, Madziva R, Mataruse N, Mbunge E, et al. Prevalence and determinants of full immunization among children under five in sub-Saharan Africa: A systematic review and meta-analysis (2013–2025). Vaccine. 2026;75:128228.
50. Demsash AW, Chereka AA, Walle AD, Kassie SY, Bekele F, Bekana T. Machine learning algorithms' application to predict childhood vaccination among children aged 12-23 months in Ethiopia: Evidence 2016 Ethiopian Demographic and Health Survey dataset. PLoS One. 2023;18(10):e0288867.
51. Mihret BG, Belay BN, Kelkay JM, Amare AM, Tarekegn AD, Asegu AA, et al. Design a model to predict incomplete immunization among Ethiopian children using ensemble machine learning algorithms. Scientific Reports. 2025;16(1):1933.
52. Endehabtu BF, Alemu K, Mengiste SA, Zelalem M, Gullslett MK, Tilahun B. Modeling Zero-Dose Children in Ethiopia: A Machine Learning Perspective on Model Performance and Predictor Variables. JMIR Pediatrics and Parenting. 2026;9.
53. Tadese ZB, Nigatu AM, Yehuala TZ, Sebastian Y. Prediction of incomplete immunization among under-five children in East Africa from recent demographic and health surveys: a machine learning approach. Scientific Reports. 2024;14(1):11529.
54. Croft T, Allen CK, Zachary BW, Arnold F, Assaf S, Balian S. Guide to DHS statistics DHS-8. Rockville, Maryland, USA: ICF. 2023.
55. Javed F, Arif M. MACHINE LEARNING ALGORITHMS FOR PREDICTIVE ANALYTICS IN BIG DATA. International Journal of Computational and Biological Sciences. 2024;1:15-36.
56. MacNell N, Feinstein L, Wilkerson J, Salo PM, Molsberry SA, Fessler MB, et al. Implementing machine learning methods with complex survey data: Lessons learned on

the impacts of accounting sampling weights in gradient boosting. PLOS ONE. 2023;18(1):e0280387.
57. Organization WH. Harmonizing vaccination coverage measures in household surveys: a primer. Geneva, Switzerland: WHO. 2019.
58. Mirza I, Lemango ET, Lindstrand A. Expanded Programme on Immunization (EPI): a legacy of 50 years and the road ahead. MDPI; 2025. p. 649.
59. Rajan A. Data Cleaning for Machine Learning Systems-A Survey. 2024.
60. Syahrizal M, Aripin S, Utomo DP, Mesran M, Sarwandi S, Hasibuan NA, editors. The application of the K-NN imputation method for handling missing values in a dataset. AIP Conference Proceedings; 2024: AIP Publishing LLC.
61. Joel LO, Doorsamy W, Paul BS. A comparative study of imputation techniques for missing values in healthcare diagnostic datasets. International Journal of Data Science and Analytics. 2025;20(7):6357-73.
62. Aracri F, Bianco MG, Quattrone A, Sarica A. Bridging the gap: Missing data imputation methods and their effect on dementia classification performance. Brain Sciences. 2025;15(6):639.
63. Santos MS, Abreu PH, Fernández A, Luengo J, Santos J. The impact of heterogeneous distance functions on missing data imputation and classification performance. Engineering Applications of Artificial Intelligence. 2022;111:104791.
64. Cubillos M, Wøhlk S, Wulff JN. A bi-objective k-nearest-neighbors-based imputation method for multilevel data. Expert Systems with Applications. 2022;204:117298.
65. Xu H, Pang G, Wang Y, Wang Y. Deep isolation forest for anomaly detection. IEEE Transactions on Knowledge and Data Engineering. 2023;35(12):12591-604.
66. Samuels J, Samuels J. One-hot encoding and two-hot encoding: an introduction. Preprint at. 2024;10.
67. Dhawas P, Dhore A, Bhagat D, Pawar RD, Kukade A, Kalbande K. Big data preprocessing, techniques, integration, transformation, normalisation, cleaning, discretization, and binning. Big Data Analytics Techniques for Market Intelligence: IGI Global Scientific Publishing; 2024. p. 159-82.
68. Cheng X. A comprehensive study of feature selection techniques in machine learning models. Available at SSRN 5154947. 2024.
69. Brownlee J. Recursive feature elimination (RFE) for feature selection in Python. Machine Learning Mastery. 2020;25.
70. Altalhan M, Algarni A, Alouane MT-H. Imbalanced data problem in machine learning: A review. IEEE Access. 2025.
71. Pradipta GA, Wardoyo R, Musdholifah A, Sanjaya INH, Ismail M, editors. SMOTE for handling imbalanced data problem: A review. 2021 sixth international conference on informatics and computing (ICIC); 2021: IEEE.
72. Elreedy D, Atiya AF. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Information Sciences. 2019;505:32-64.
73. Wongvorachan T, He S, Bulut O. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. Information. 2023;14(1):54.
74. Yang B, Gül M, Chen Y. Comparative analysis of deep learning and tree-based models in power demand prediction: Accuracy, interpretability, and computational efficiency. Journal of Building Physics. 2025;49(1):127-69.

75. Kundu LR, Majumder AK. Hyperparameter-Tuned Machine Learning Techniques for the Prediction of Diabetes Status Among Bangladeshi Adults. Advances in Public Health. 2026;2026(1):8525279.
76. Mondal S, Maity R, Nag A, Ghosh S. Fetal health risk prediction using ensemble-based machine learning approaches. Knowledge and Information Systems. 2025:1-35.
77. Theng D, Bhoyar KKJK, Systems I. Feature selection techniques for machine learning: a survey of more than two decades of research. 2024;66(3):1575-637.
78. Büyükkeçeci M, Okur MCJGUJoS. A comprehensive review of feature selection and feature selection stability in machine learning. 2023;36(4):1506-20.
79. Barzani AR, Pahlavani P, Ghorbanzadeh O, Gholamnia K, Ghamisi PJF. Evaluating the impact of recursive feature elimination on machine learning models for predicting forest fire-prone zones. 2024;7(12):440.
80. Bulut O, Tan B, Mazzullo E, Syed A. Benchmarking Variants of Recursive Feature Elimination: Insights from Predictive Tasks in Education and Healthcare. Information. 2025;16(6):476.
81. Jason B. Recursive feature elimination (RFE) for feature selection in Python. MachineLearningMastery Com. 2020;27.
82. Wah YB, Ismail A, Azid N, Niswah N, Jaafar J, Aziz IA, et al. Machine Learning and Synthetic Minority Oversampling Techniques for Imbalanced Data: Improving Machine Failure Prediction. Computers, Materials & Continua. 2023;75(3).
83. Hairani H, Widiyaningtyas T, Prasetya DDJJIJoIV. Addressing class imbalance of health data: a systematic literature review on modified synthetic minority oversampling technique (SMOTE) strategies. 2024;8(3):1310-8.
84. Owusu-Adjei M, Ben Hayfron-Acquah J, Frimpong T, Abdul-Salaam GJPDH. Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. 2023;2(11):e0000290.
85. Hancock JT, Khoshgoftaar TM, Liang Q. A problem-agnostic approach to feature selection and analysis using SHAP. Journal of Big Data. 2025;12(1):12.
86. Siargkas A, Tsakiridis I, Kappou D, Mamopoulos A, Papastefanou I, Dagklis T. Machine Learning Models for the Prediction of Preterm Birth at Mid-Gestation Using Individual Characteristics and Biophysical Markers: A Cohort Study. Children. 2025;12(11):1451.
87. Nantongo BA, Nabukenya J, Nabende P, Kamulegeya J. A retrospective cohort study on predicting infants at a risk of defaulting routine immunization in Uganda using machine learning models. JAMIA open. 2024;7(4):ooae132.
88. Asnake AA, Seifu BL, Gebrehana AK. Prediction of zero-dose children using supervised machine learning algorithm in Tanzania: evidence from the recent 2022 Tanzania Demographic and Health Survey. BMJ open. 2025;15(3):e097395.
89. Krishnamoorthy Y, Rehman T. Impact of antenatal care visits on childhood immunization: a propensity score-matched analysis using nationally representative survey. Family Practice. 2022;39(4):603-9.
90. Agbi FA, Lulin Z, Asamoah EO. Quality of communication between healthcare providers and pregnant women: impact on maternal satisfaction, health outcomes, and shared decision-making. Trends J Sci Res. 2023;2(1):3-10.

91. Bashir S, Ansari AH, Sultana A. Knowledge, attitude, and practice on antenatal care among pregnant women and its association with sociodemographic factors: a hospital-based study. Journal of patient experience. 2023;10:23743735231183578.
92. Control CfD, Prevention. Child and adolescent immunization schedule by age. Recommendations for ages. 2023;18.
93. Krishnamoorthy Y, Rehman T. Impact of antenatal care visits on childhood immunization: a propensity score-matched analysis using nationally representative survey. Family Practice. 2021;39:603-9.
94. Tiruneh GT, Chitashvili T, Hunegnaw BM, Shewangzaw T, Bogale B, Fesseha N, et al. Effective coverage of maternal and newborn health services across the antepartum and peripartum continuum in primary health care in Ethiopia. Reproductive Health. 2025;23(1):14.
95. Chowdhury AI, Habib MA, Rahman T. Factors Influencing the Utilization of Antenatal Care, Institutional Delivery, and Postnatal Care Services Among Women in Bangladesh. Makara Journal of Health Research. 2022;26:146-52.
96. Tilahun WM, Geremew H, Yohannes Abay L, Kebie AB, Simegn MB. Determinants of tetanus immunization among pregnant women where tetanus has not been eliminated: a multilevel analysis of 6 countries. Frontiers in global women's health. 2025;6:1481771.
97. Gashaw A. Determinants of TT vaccine completion among mothers who gave birth with in the past six months in Gozamn District, Ethiopia. International Journal of Africa Nursing Sciences. 2026;24:100969.
98. Jiao B, Iversen I, Sato R, Pecenka C, Khan S, Baral R, et al. Association between achieving adequate antenatal care and health-seeking behaviors: A study of Demographic and Health Surveys in 47 low-and middle-income countries. PLoS medicine. 2024;21(7):e1004421.
99. Khan T, Malik S, Rafeekh L, Halder S, Desai S, Das Bhattacharya S. Facilitators and barriers to maternal immunization and strategies to improve uptake in low-income and lower-middle income countries: a systematic review. Human Vaccines & Immunotherapeutics. 2024;20(1):2411823.
100. Zemariam AB, Abebe GK, Kassa MA, Alamaw AW, Molla RW, Abate BB, et al. Immunization coverage and its associated factors among children aged 12-23 months in Ethiopia: An umbrella review of systematic review and meta-analysis studies. 2024;19(3):e0299384.
101. Nour TY, Farah AM, Ali OM, Osman MO, Aden MA, Abate KH. Predictors of immunization coverage among 12–23 month old children in Ethiopia: systematic review and meta-analysis. BMC Public Health. 2020;20(1):1803.
102. Salah F, Tafere C, Ejigu LJ, Tadesse F, Gedif A, Mulatu S. Vaccines coverage and associated factors among children aged 12-23 months in the Pawie district, Ethiopia: A cross-sectional study. Journal of virus eradication. 2024;10(3):100391.
103. Zemariam AB, Abebe GK, Kassa MA, Alamaw AW, Molla RW, Abate BB, et al. Immunization coverage and its associated factors among children aged 12–23 months in Ethiopia: An umbrella review of systematic review and meta-analysis studies. PLOS ONE. 2024;19(3):e0299384.
104. Tesfa GA, Yehualashet DE, Getnet A, Bimer KB, Seboka BT. Spatial distribution of complete basic childhood vaccination and associated factors among children aged 12–23 months in Ethiopia. A spatial and multilevel analysis. PLOS ONE. 2023;18(1):e0279399.

105. Allan S, Adetifa IMO, Abbas K. Inequities in childhood immunisation coverage associated with socioeconomic, geographic, maternal, child, and place of birth characteristics in Kenya. BMC Infectious Diseases. 2021;21(1):553.
106. Okello G, Izudi J, Ampeire I, Nghania F, Dochez C, Hens N. Two decades of regional trends in vaccination completion and coverage among children aged 12-23 months: an analysis of the Uganda Demographic Health Survey data from 1995 to 2016. BMC Health Services Research. 2022;22(1):40.
107. Aalemi AK, Shahpar K, Mubarak MY. Factors influencing vaccination coverage among children age 12-23 months in Afghanistan: Analysis of the 2015 Demographic and Health Survey. PLoS One. 2020;15(8):e0236955.
108. Kumar R, Fazal K, Choudhary AN. Assessment Of Primary Immunization Coverage Among The Children Aged 12-23 Months In Urban Area Of Bhagalpur City. Journal of Pharmaceutical Negative Results. 2023;14(2).
109. Schön M, Heesemann E, Ebert C, Subramanyam M, Vollmer S, Horn S. How to ensure full vaccination? The association of institutional delivery and timely postnatal care with childhood vaccination in a cross-sectional study in rural Bihar, India. PLOS Global Public Health. 2022;2(5):e0000411.
110. Fatinah Y, Qolbina M, Alfaqeeh M, Insani WN, Sinuraya RK, Hamad A, et al. Childhood immunization barriers and interventions in low-and middle-income countries: A scoping review of access, delivery systems, and health literacy. Human Vaccines & Immunotherapeutics. 2026;22(1):2651508.
111. Aynalem BY, Zerihun T, Alemu AA, Meselu BT, Zeleke LB, Kassa GM. Home delivery and associated factors among women who gave birth after antenatal care follow-up in the last 6 months in Jabitehnan District, Northwest Ethiopia: mixed community-based study. Frontiers in global women's health. 2024;5:1458453.
112. Zenbaba D, Sahiledengle B, Dibaba D, Bonsa M. Utilization of Health Facility-Based Delivery Service Among Mothers in Gindhir District, Southeast Ethiopia: A Community-Based Cross-Sectional Study. Inquiry : a journal of medical care organization, provision and financing. 2021;58:469580211056061.
113. Schön M, Heesemann E, Ebert C. How to ensure full vaccination? The association of institutional delivery and timely postnatal care with childhood vaccination in a cross-sectional study in rural Bihar, India. 2022;2(5):e0000411.
114. Wojcieszek AM, Bonet M, Portela A, Althabe F, Bahl R, Chowdhary N, et al. WHO recommendations on maternal and newborn care for a positive postnatal experience: strengthening the maternal and newborn care continuum. BMJ Global Health. 2023;8(Suppl 2):e010992.
115. Wojcieszek AM, Bonet M. WHO recommendations on maternal and newborn care for a positive postnatal experience: strengthening the maternal and newborn care continuum. 2023;8(Suppl 2).
116. Hodgson A, Jimba M. PLoS Med.
117. Mekonnen ZA, Gelaye KA, Were MC, Tilahun B. Timely completion of vaccination and its determinants among children in northwest, Ethiopia: a multilevel analysis. BMC Public Health. 2020;20(1):908.
118. Tamirat KS, Sisay MM. Full immunization coverage and its associated factors among children aged 12–23 months in Ethiopia: further analysis from the 2016 Ethiopia demographic and health survey. BMC Public Health. 2019;19(1):1019.

119. Debie A, Lakew AM, Tamirat KS, Amare G, Tesema GA. Complete vaccination service utilization inequalities among children aged 12-23 months in Ethiopia: a multivariate decomposition analyses. Int J Equity Health. 2020;19(1):65.
120. Geweniger A, Abbas KM. Childhood vaccination coverage and equity impact in Ethiopia by socioeconomic, geographic, maternal, and child characteristics. Vaccine. 2020;38(20):3627-38.
121. Shiferie F, Gebremedhin S, Andargie G, Tsegaye DA, Alemayehu WA, Fenta TG. Decomposition Analysis of Socioeconomic Inequalities in Vaccination Dropout in Remote and Underserved Settings in Ethiopia. The American journal of tropical medicine and hygiene. 2024;111(1):196-204.
122. Memirie ST, Tolla MT, Rumpler E, Sato R, Bolongaita S, Tefera YL, et al. Out-of-pocket expenditures and financial risks associated with treatment of vaccine-preventable diseases in Ethiopia: A cross-sectional costing analysis. PLOS Medicine. 2023;20(3):e1004198.
123. Nigatu T, Abraham L, Willems H, Tilaye M, Tiruneh F, Gebru F, et al. The status of immunization program and challenges in Ethiopia: A mixed method study. SAGE Open Medicine. 2024;12:20503121241237115.
124. Salako J, Bakare D, Uchendu OC, Bakare AA, Graham H, Falade AG. Factors associated with immunization status among children aged 12-59 months in Lagelu local government area, Ibadan: a cross-sectional study. The Pan African medical journal. 2024;47:35.
125. Sibeudu FT, Uzochukwu BS, Onwujekwe OE. Rural–urban comparison of routine immunization utilization and its determinants in communities in Anambra states, Nigeria. SAGE open medicine. 2019;7:2050312118823893.
126. Sako S, Gilano G, Hailegebreal S. Determinants of childhood vaccination among children aged 12–23 months in Ethiopia: a community-based cross-sectional study. BMJ Open. 2023;13(3):e069278.
127. Tamirat KS, Sisay MM. Full immunization coverage and its associated factors among children aged 12–23 months in Ethiopia: further analysis from the 2016 Ethiopia demographic and health survey. BMC public health. 2019;19(1):1019.
128. Nour TY, Farah AM, Ali OM, Osman MO, Aden MA, Abate KH. Predictors of immunization coverage among 12–23 month old children in Ethiopia: systematic review and meta-analysis. BMC public health. 2020;20(1):1803.
129. Budu E, Darteh EKM, Ahinkorah BO, Seidu A-A, Dickson KS. Trend and determinants of complete vaccination coverage among children aged 12-23 months in Ghana: analysis of data from the 1998 to 2014 Ghana demographic and health surveys. Plos one. 2020;15(10):e0239754.
130. Ozigbu CE, Olatosi B, Li Z, Hardin JW, Hair NL. Correlates of zero-dose vaccination status among children aged 12–59 months in Sub-saharan Africa: a multilevel analysis of individual and contextual factors. Vaccines. 2022;10(7):1052.
131. Shital S, Shruti D, Gautam D, Swami A. Impact of Maternal and Child Health (MCH) Programs on Immunization Coverage in Rural and Urban Areas. 2025.
132. Asmare G, Madalicho M, Sorsa A. Disparities in full immunization coverage among urban and rural children aged 12-23 months in southwest Ethiopia: A comparative cross-sectional study. Human Vaccines & Immunotherapeutics. 2022;18(6):2101316.
133. Santos TM, Cata-Preta BO. Exploring the "Urban Advantage" in Access to Immunization Services: A Comparison of Zero-Dose Prevalence Between Rural, and Poor

and Non-poor Urban Households Across 97 Low- and Middle-Income Countries. 2024;101(3):638-47.
134. Tebeje TM, Aregu MB, Asgedom YS. From poverty to health: intraurban inequalities in child health indicators in low-income and middle-income countries during the SDG era. 2025;10(8).
135. Manoli A, Chandrabose GS, Dhakishnamoorthy A, Behea KK, Surti S, Patil RV, et al. Community-Based Strategies for Improving Vaccine Uptake Among Rural Populations. Genetics and Molecular Research. 2025;24(2):1-6.
136. Jil JPM. Immunization status of children aged 12-23 months in Jonglei State, South Sudan: a cross-sectional epidemiologic study. Pan African Medical Journal. 2022;41(1).
137. Gebretnsae H, Ayele B. Vaccination status and its determinants among children aged 12-23 months in Tigray, northern Ethiopia: A zero-inflated Poisson regression analysis. 2025;20(7):e0327854.
138. Barrow A, Afape AO, Cham D, Azubuike PC. Uptake and determinants of childhood vaccination status among children aged 0–12 months in three West African countries. BMC Public Health. 2023;23(1):1093.
139. Fagbamigbe AF, Lawal TV, Atoloye KA. Evaluating the performance of different Bayesian count models in modelling childhood vaccine uptake among children aged 12–23 months in Nigeria. BMC Public Health. 2023;23(1):1197.
140. Tiruneh GT, Hunegnaw BM, Enkubahiri S, Bogale B, Shewangzaw T, Fesseha N, et al. Associations between integrated family planning, maternal and newborn health, and immunization services and adoption of postpartum family planning and immunization services in Ethiopia. BMC Public Health. 2025;25(1):3258.
141. Girma Tareke K, Feyissa GT, Kebede Y. Exploration of barriers to postnatal care service utilization in Debre Libanos District, Ethiopia: A descriptive qualitative study. Frontiers in global women's health. 2022;Volume 3 - 2022.
142. Mespreuve AS, Apers L, Moller AB. Postnatal quality of care measures for mothers and newborns at home: A scoping review. 2024;4(8):e0003498.
143. Altraifi AA, Albasheer O. Postnatal care utilisation and health beliefs among mothers in the Jazan region of Saudi Arabia: a cross-sectional study. 2025;15(8):e088255.
144. Asumah MN, Abubakari A, Abdulai AM, Nukpezah RN, Adomako-Boateng F, Faridu A-W, et al. Sociodemographic and maternal determinants of postnatal care utilization: a cross-sectional study. SAGE Open Nursing. 2023;9:23779608231206759.
145. Amsalu G, Talie A, Gezimu W. Non-utilization of postnatal care and its associated factors among women who gave birth in rural districts of Northern Ethiopia: A community-based mixed-method study. 2022;18:17455057221125091.
146. Planning HIPiF. Family planning and immunization integration: reaching postpartum women with family planning services. USAID Washington, DC; 2021.
147. Debie A, Lakew AM, Tamirat KS, Amare G, Tesema GA. Complete vaccination service utilization inequalities among children aged 12–23 months in Ethiopia: a multivariate decomposition analyses. International Journal for Equity in Health. 2020;19(1):65.
148. Endalamew SG, Gebeye E, Assefa SK. Immunization Coverage and Its Determinants Among Children Aged 12-23 Months in East Africa: A Bayesian Hierarchical Modeling Approach Based on DHS 2019-2022. 2025;2025:6645541.

149. Machmud PB, Glasauer S, Gayatri D, Mikolajczyk R. Mother's media use and children's vaccination status in Indonesia: A community-based cross-sectional study. Global Pediatric Health. 2022;9:2333794X221092740.
150. Jejaw M, Demissie KA, Teshale G, Geberu DM, Tiruneh MG, Hagos A, et al. Estimated effect of media use on mothers' vaccination of their children in Sub-Saharan Africa: a quasi-experimental propensity score matching analysis using DHS data. BMC Public Health. 2025;25(1):2018.
151. Nazari A, Ataei R, Heydarifard Z, Mousavi A. Social media-based interventions for improving vaccine uptake, reducing hesitancy, and combating misinformation: a comprehensive systematic review and meta-analysis of RCT. BMC Public Health. 2026.
152. Oyefara J. Mothers' Media Exposure and Children Immunization Coverage in Lagos the Metropolis, Nigeria. Dutse Journal of Humanities and Social Sciences. 2014;1:313-30.
153. Leonforte F, Nicosia V, Comite P, Morlino G, Mistretta A. Media Exposure and Its Association With Vaccine Attitudes, Intentions, and Hesitancy: Systematic Review. Journal of medical Internet research. 2026;28:e74280.
154. Njenga GG, Mogambi H, Oriaso S. Media and old age: health Information-Seeking among elderly slum dwellers in Kenya. J Promkes Indones J Health Promot Health Educ. 2024;12:170-85.
155. Li L, Wood CE, Kostkova P. Vaccine hesitancy and behavior change theory-based social media interventions: a systematic review. Translational behavioral medicine. 2022;12(2):243-72.
156. Limaye RJ, Holroyd TA, Blunt M, Jamison AF, Sauer M, Weeks R, et al. Social media strategies to affect vaccine acceptance: a systematic literature review. Expert review of vaccines. 2021;20(8):959-73.
157. Johri M, Chandra D, Kone KG, Sylvestre M-P, Mathur AK, Harper S, et al. Social and behavior change communication interventions delivered face-to-face and by a mobile phone to strengthen vaccination uptake and improve child health in rural India: randomized pilot study. JMIR mHealth and uHealth. 2020;8(9):e20356.
158. Alhassan M, Akpede K. Assessment of Use of Social Media Platforms for Integrated Immunization Enlightenment by the Nasarawa State Government, Nigeria. International Journal of Sub-Saharan African Research. 2025;3(2):236-49.
159. Girmay A, Dadi AF. Full Immunization Coverage and Associated Factors among Children Aged 12-23 Months in a Hard-to-Reach Areas of Ethiopia. 2019;2019:1924941.
160. Kibreab F, Lewycka S, Tewelde A. Impact of mother's education on full immunization of children aged 12–23 months in Eritrea: population and health survey 2010 data analysis. BMC Public Health. 2020;20(1):267.
161. Jama AA. Determinants of Complete Immunization Coverage among Children Aged 11-24 Months in Somalia. International journal of pediatrics. 2020;2020(1):5827074.
162. Sani J, Halane S. Regional Disparities and Maternal Sociodemographic Determinants of Full Immunization Coverage Among Children Aged 12-23 Months in Nigeria: Insights from NDHS 2018. 2025;16:157-70.
163. Tesema GA, Tessema ZT, Tamirat KS, Teshale AB. Complete basic childhood vaccination and associated factors among children aged 12–23 months in East Africa: a multilevel analysis of recent demographic and health surveys. BMC Public Health. 2020;20(1):1837.

164. Lael-Monfared E, Sabouri S, Rahmani N, Peyman N. Health literacy, awareness and referral patterns of mothers in childhood vaccination: a cross-sectional study in Iran. BMC Public Health. 2025;26(1):325.
165. Egwuaba E, Sunday B, Sunday U. Maternal Health Literacy and Childhood Immunization Uptake: A Social and Educational Analysis in North-Central Nigeria. African Journal of Humanities and Contemporary Education Research. 2026;22:31-51.
166. Gebreyesus A, Tesfay K. Effect of maternal education on completing childhood vaccination in Ethiopia: systematic review and meta-analysis. Sci Rep. 2024;14(1):17453.
167. Rahmadani RA, Akib A, Barr JA. Maternal factors influencing complete basic immunization: A Scoping Review. Jurnal Ilmiah Kesehatan Sandi Husada. 2025;14(2):284-95.
168. Lee HY, Oh J. Association between maternal literacy and child vaccination in Ethiopia and southeastern India and the moderating role of health workers: a multilevel regression analysis of the Young Lives study. 2019;12(1):1581467.
169. Meghwal K, Kumar P. Impact of maternal education on knowledge and adherence to immunization schedules: A systematic review. Journal of Neonatal Surgery. 2025;14(26S):264-71.
170. Vhawal V, RJ K, Kanase N, editors. THE ROLE OF MATERNAL EDUCATION IN ENHANCING CHILD HEALTH: FOCUS ON VACCINATION AND NUTRITIONAL STATUS. Obstetrics & Gynaecology Forum; 2024.
171. Noh J-W, Kim Y-m, Akram N, Yoo K-B, Park J, Cheon J, et al. Factors affecting complete and timely childhood immunization coverage in Sindh, Pakistan; A secondary analysis of cross-sectional survey data. PloS one. 2018;13(10):e0206766.
172. Hayir TMM, Magan MA, Mohamed LM, Mohamud MA, Muse AA. Barriers for full immunization coverage among under 5 years children in Mogadishu, Somalia. Journal of family medicine and primary care. 2020;9(6):2664-9.
173. Musuka G, Moyo E, Mano O, Madziva R, Gashema P, Iradukunda P, et al. Paternal factors affecting under-five immunization status in Sub-Saharan Africa: A systematic review and meta-analysis. Vaccine. 2025;64:11.
174. Azubuike CD, Ardelt M. The influence of paternal characteristics on childhood vaccination in Nigeria. Discover Public Health. 2025;22(1):123.
175. UNICEF. Strengthening Immunization for Children: Tackling Gender Barriers and Promoting Parenting Alliance. published; 2024.
176. Abegaz MY, Seid A, Awol SM, Hassen SL. Determinants of incomplete child vaccination among mothers of children aged 12-23 months in Worebabo district, Ethiopia: Unmatched case-control study. PLOS Glob Public Health. 2023;3(8):e0002088.
177. Worku BT, Wordofa EA, Senbeto G, Zinab B, Kebede EB. Age-appropriate vaccination and associated factors among children aged 12-35 months in Ethiopia: A multi-level analysis. 2024;19(10):e0305801.
178. Abamecha F, Debela GT, Birhanu N, Siraneh Y, Tamiru D. PLoS One.
179. Nadella P, Smith ER, Muhihi A, Noor RA, Masanja H, Fawzi WW, et al. Determinants of delayed or incomplete diphtheria-tetanus-pertussis vaccination in parallel urban and rural birth cohorts of 30,956 infants in Tanzania. BMC Infectious Diseases. 2019;19(1):188.
180. Adesina MA, Olufadewa II, Oladele RI, Solagbade A, Olaoyo C. Determinants of childhood immunization among rural mothers in Nigeria. Population Medicine. 2023;5(September):1-7.

181. Nozaki I, Hachiya M, Kitamura T. Factors influencing basic vaccination coverage in Myanmar: secondary analysis of 2015 Myanmar demographic and health survey data. BMC public health. 2019;19(1):242.
182. Sarker AR, Akram R, Ali N, Chowdhury ZI, Sultana M. Coverage and determinants of full immunization: vaccination coverage among Senegalese children. Medicina. 2019;55(8):480.
183. Hatijar, Rahmadani R, Akib A, Barr J. Maternal factors influencing complete basic immunization: A Scoping Review. Jurnal Ilmiah Kesehatan Sandi Husada. 2025;14:284-95.
184. Balogun FM, Omobowale OC. "Grandmothers as gatekeepers": The role of older women in addressing vaccine hesitancy in urban slum communities in Ibadan, Nigeria. PLOS Global Public Health. 2026;6(1):e0005786.
185. Yorulmaz-Demir DS, Kocoglu-Tanyer D. The Effect of the Vaccine Education Program Prepared Using The Integrated Change Model on Mothers' Vaccine Advocacy and Vaccine Literacy: A Waitinglist Randomized Controlled Trial. Journal of pediatric health care : official publication of National Association of Pediatric Nurse Associates & Practitioners. 2025;39(3):424-35.
186. Mathewos Oridanigo E, Kassa B. Utilization of Skilled Birth Attendance among Mothers Who Gave Birth in the Last 12 Months in Kembata Tembaro Zone. Advances in medicine. 2022;2022:8180387.
187. Ahinkorah BO, Seidu A-A, Agbaglo E, Adu C, Budu E, Hagan JE, et al. Determinants of antenatal care and skilled birth attendance services utilization among childbearing women in Guinea: evidence from the 2018 Guinea Demographic and Health Survey data. BMC Pregnancy and Childbirth. 2021;21(1):2.
188. Ameyaw EK. Skilled birth attendance: safeguarding mothers and newborns. The Lancet Global Health. 2024;12(7):e1071-e2.