

Determinants of Skilled Birth Attendant Utilization in Bangladesh: A Cross-Sectional Study of BDHS 2022 Using Statistical and Machine Learning Methods

Md. Rashidul Islam Rifat

rashidulislamrifat@gmail.com

Mawlana Bhashani Science and Technology University

Md. Deluar Jahan Moley

Mawlana Bhashani Science and Technology University

Arzo Ahmed

Mawlana Bhashani Science and Technology University

Md. Mizanur Rahman

Mawlana Bhashani Science and Technology University

Most.Farjana Yesmin

Mawlana Bhashani Science and Technology University

Md. Onthor

Mawlana Bhashani Science and Technology University

Tasnia Rahman Shefa

Mawlana Bhashani Science and Technology University

Research Article

Keywords: Skilled birth attendant, Bangladesh, BDHS 2022, maternal health, machine learning, logistic regression, health inequality

Posted Date: May 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9480632/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Skilled birth attendant (SBA) utilization is a key strategy for reducing maternal and neonatal mortality. Despite notable progress in Bangladesh, substantial socioeconomic and regional disparities persist. This study aimed to identify the key determinants of SBA utilization using a hybrid analytical approach that integrates statistical and machine learning methods.

Methods

Data from 4,646 ever-married women were obtained from the Bangladesh Demographic and Health Survey (BDHS) 2022. A hybrid feature selection framework combining chi-square tests, LASSO regression, Boruta algorithm, and variance inflation factor (VIF) analysis was applied. Seven machine learning models were developed, with class imbalance addressed using the Synthetic Minority Over-sampling Technique (SMOTE). Model performance was evaluated using accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve (AUC), with robustness assessed through stratified k-fold cross-validation. Logistic regression and SHAP analysis were used to enhance interpretability.

Results

Overall, 72.9% of women utilized SBA services. Among the models, XGBoost demonstrated the best overall predictive performance (accuracy = 0.75, sensitivity = 0.91, F1-score = 0.84), while logistic regression and support vector machine (SVM) showed the highest discriminative ability (AUC = 0.77). Key factors associated with SBA utilization included wealth index, maternal education, antenatal care (ANC) utilization, geographic region, media exposure, religion, and age at first birth. Logistic regression analysis revealed higher odds of SBA utilization among women in Khulna division (OR = 4.11), non-Muslim women (OR = 2.86), and those with media exposure (OR = 1.39), whereas inadequate ANC visits, higher parity, and lower socioeconomic status were associated with reduced utilization.

Conclusion

SBA utilization in Bangladesh is associated with socioeconomic inequality, healthcare access, and regional disparities. The proposed hybrid analytical framework improves both predictive performance and interpretability, offering a practical and scalable approach for evidence-based maternal health policymaking in Bangladesh and similar low and middle-income settings.

Background

Maternal mortality remains a major global public health challenge, reflecting profound inequalities in access to quality maternal healthcare. Each year, it is estimated that 287,000 women die due to pregnancy- and childbirth-related complications, with most of the deaths occurring in low- and middle-income countries (1). The concept of evidence-based maternity care during birth is one of the cornerstone approaches for reducing maternal and neonatal mortality and is enshrined in the Sustainable Development Goal (SDG) 3.1, which aims to reduce maternal mortality rates to less than 70 deaths per 100,000 live births by 2030 (1, 2). According to the WHO and its partner agencies, a skilled birth attendant (SBA) is a midwife, nurse, or physician who is trained to carry out normal deliveries and identify complications and referrals (3). Scholarly evidence shows that competent birth attendance significantly lowers maternal and perinatal death rates (4, 5).

Skilled birth attendance (SBA) has increased worldwide, but access is still unequal for women in poorer or hard-to-reach communities, who are less likely to receive skilled care (6, 7). The greatest barriers to access to skilled birth attendance (SBA) in low- and middle-income settings are geography and socioeconomic disadvantage, with rural, poor, and less-educated women particularly underserved (8). The case of Bangladesh represents both improvements and difficulties. By 2020, the country had achieved a maternal mortality ratio of approximately 160–165 deaths per 100,000 live births, which is notably better than the case at the beginning of the 1990s and has been realized through the growth of primary health care, community-based interventions and higher rates of antenatal care (ANC) utilization (2, 9). However, findings from the Bangladesh Demographic and Health Survey (BDHS) 2022 indicate that the utilization of skilled birth attendants varies substantially across population subgroups, with higher coverage observed among urban residents and wealthier households (10, 6).

These urban, rural and social economic differences highlight the inequities of the system in maternal care. The literature lists maternal education, household wealth, exposure to the media, and antenatal care utilization as important predictors of SBA utilization in Bangladesh (11–13). Other determinants include the age of the mother, parity, religion, subdivision of regions, and distance to well-equipped health facilities (Islam *et al.*, 2021; Kamal, 2021). For example, women with secondary or higher education are more likely to give birth under skilled supervision; when they are in the lowest wealth quintile, the probability is much lower (13). Similarly, institutional delivery is positively associated with earlier initiation and more frequent ANC visits (14).

Regardless of the strength of these associations, Bangladeshi studies have predominantly used logistic regression and multilevel logistic models to analyze maternal healthcare utilization (12, 15, 11, 16). Such approaches often overlook nonlinearities and high-dimensional relationships among socioeconomic, demographic, and health-service variables, although they can be interpreted. The inherently complex and multifaceted nature of determinants of SBA utilization has led to the development of sophisticated computational tools, specifically machine learning (ML), as powerful predictive modeling and feature detection tools (17, 18).

Machine learning has emerged as a powerful approach in the research of global health, as it allows the prediction of health outcomes and the identification of patterns of latent risk that are not always apparent in traditional analytical frameworks. Machine learning has been used to anticipate preterm birth, low birthweight, and obstetric complications of maternal health (17, 19). Random forest, XGBoost, and support vector machines are often compared with traditional methods and outperform them not only in low- and middle-income countries but also in cases where survey and clinical data are combined (19). In Bangladesh, a recent study that focused on region-specific features and used BDHS 2022 data demonstrated accurate prediction of low birthweight rates, which highlights the importance of data-driven analytics in policy-making processes (19). The existing international data suggest that ensemble ML methods are often more effective than conventional regression methods in capturing nonlinear interactions among sociodemographic factors (17).

However, the incorporation of ML into health analytics is still limited, especially when studies combine conventional statistical inference with ML-based predictive modeling to examine the utilization of SBAs in South Asia. The methodologies employed in the Bangladeshi literature are mostly inferential with limited emphasis on predictive modeling and feature importance analysis (11, 20). Combining these methodological traditions can result in better prediction performance and maintain interpretability, thereby generating findings that are more relevant for evidence-based policymaking. Moreover, the sharp regional and socioeconomic divisions of Bangladesh, such as disparities between urban centers like Dhaka and rural regions across divisions, make the country an ideal setting for applying hybrid analytical methods to unravel the key determinants of the SBA utilization (6, 20). Furthermore, existing studies rarely integrate rigorous feature selection techniques with machine learning models while simultaneously ensuring interpretability, which limits their applicability for policy decision-making.

To address these gaps, the present study adopts an integrated analytical framework that combines conventional statistical modeling with advanced machine learning techniques to comprehensively examine the determinants of skilled birth attendant (SBA) utilization in Bangladesh. Unlike prior studies that primarily rely on inferential approaches, this study simultaneously emphasizes both explanatory and predictive modeling to capture complex, nonlinear relationships among sociodemographic, healthcare, and geographic factors. Using nationally representative BDHS 2022 data, this study aims to (i) identify the key predictors of SBA utilization through robust feature selection techniques, (ii) evaluate and compare the predictive performance of multiple machine learning algorithms, and (iii) quantify the strength and direction of associations using logistic regression for policy-relevant interpretation. By integrating interpretable statistical inference with high-performing machine learning models, this study provides a more comprehensive understanding of SBA utilization and offers an evidence-based framework to support targeted maternal health interventions and data-driven policymaking in Bangladesh and similar low- and middle-income settings.

This study contributes to the literature by integrating interpretable statistical inference with high-performing machine learning models, thereby providing actionable insights for improving maternal healthcare utilization.

Methods

Study Design and Data Source

The analysis was based on the 2022 Bangladesh Demographic and Health Survey (BDHS), which is the ninth nationally representative survey that outlines the demographic and health-related conditions of Bangladeshi women and children (21). The BDHS provides comprehensive data on fertility, maternal and child health, family planning, noncommunicable diseases, mental health and access to healthcare. NIPORT carried out the survey in cooperation with the ICF and the USAID and within the framework of the DHS Program (22).

Two-stage stratified cluster sampling was employed by the BDHS and incorporated 675 enumeration areas (EAs) and 45 households in each EA (21). The official 2022 BDHS report contains detailed methodological descriptions and procedures used in surveys.

Study Population and Sample

The study population comprised ever-married women aged 15–49 years who participated in the Bangladesh Demographic and Health Survey (BDHS) 2022. A total of 4,646 women with complete information on skilled birth attendant (SBA) utilization and all relevant covariates were included in the final analysis.

A complete case analysis approach was applied, whereby observations with missing values in the outcome or any explanatory variables were excluded.

Outcome Variable

The use of a skilled birth attendant (SBA) during birth was the key outcome variable. On the basis of the definitions of the World Health Organization (WHO) and the BDHS, SBAs are qualified doctors, nurses, midwives, paramedics, family welfare visitors (FWVs), community skilled birth attendants (CSBAs), and sub-assistant community medical officers (SACMOs) (3, 21). All other attendants, including traditional birth attendants, untrained providers, relatives and neighbors, were categorized as unskilled. For analysis, the number of births attended by SBAs was coded as 1, and the number of births attended by unskilled personnel was coded as 0. All of the live births and stillbirths in the two years before the survey were included in the analysis.

Predictor Variables and Feature Selection

A comprehensive set of categorical predictor variables related to demographic, socioeconomic, and healthcare factors was initially considered. First, chi-square tests of independence were performed to assess the association between each categorical predictor and the outcome variable (SBA utilization) (23). Variables with statistically significant associations ($p < 0.05$) were retained, allowing initial dimensionality reduction and removal of irrelevant predictors.

The selected variables were then transformed into numerical format using one-hot encoding, with the first category dropped to avoid the dummy variable trap and reduce multicollinearity. The dataset was subsequently divided into training and testing subsets using an 80:20 stratified split to preserve class distribution (24).

To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied exclusively to the training dataset to prevent data leakage (25). Feature selection was then performed on the balanced training data using a hybrid approach combining LASSO (Least Absolute Shrinkage and Selection Operator) regression and the Boruta algorithm. LASSO was used to identify features with non-zero coefficients (26), while Boruta selected relevant features based on Random Forest importance scores. The final feature set was derived from the union of both methods, ensuring identification of both linear and nonlinear relationships.

To further ensure model stability and eliminate multicollinearity, an iterative Variance Inflation Factor (VIF) analysis was conducted on the selected features. Variables with VIF values greater than 5 were sequentially removed until all remaining predictors satisfied the threshold ($VIF < 5$) (27).

The final set of predictors used for modeling is presented in Table 1.

Table 1
Final selected predictors after hybrid feature selection and VIF analysis

Variable	Categories	VIF Value
Right age at marriage	Right time	3.32
Age	25–29	3.31
Education level	Primary	3.25
Right age at marriage	Later	3.17
Age	30–34	3.15
Education level	Secondary	3.04
Number of living children	≥ 3 children	2.94
Age	20–24	2.41
Division	Chattogram	2.21
Wealth index	Poor	2.20
Wealth index	Rich	2.19
Age	35–39	2.17
Division	Dhaka	2.06
Husband occupation	Hard work	1.99
Division	Mymensingh	1.98
Husband occupation	Farmer	1.98
Number of living children	2 children	1.92
Division	Sylhet	1.87
Division	Rangpur	1.86
Division	Rajshahi	1.73
Division	Khulna	1.72
Antenatal care visits	Once	1.71
Education level	No education	1.67
Antenatal care visits	Two times	1.56
Age at first birth	≤ 20 years	1.55
Husband occupation	Service holder	1.42
Use of internet	Yes	1.38
Timing of first ANC check	Within first trimester	1.34
Media exposure	Yes	1.33
Antenatal care visits	Three times	1.30
Relationship to household head	Wife	1.27
Type of residence	Urban	1.27
Heard family planning	Yes	1.18
Type of toilet facility	Unhygienic	1.15
Husband occupation	Unemployed	1.12

Variable	Categories	VIF Value
Age	45–49	1.09
Religion	Non-Muslim	1.06
Antenatal care provider	Unskilled provider	1.05

Software and libraries

All statistical analyses and machine learning models were implemented using Python (version 3.10, Anaconda Inc., USA). The pandas (28) and NumPy libraries were used to preprocess and manipulate the data (29). SciPy was used to conduct statistical tests, including chi-square tests (23). Scikit-learn was used to develop machine learning models, including logistic regression, decision tree, random forest, k-nearest neighbors, support vector machine (SVM), and naive Bayes (30), and XGBoost was used to develop gradient boosting (31). Matplotlib (32) and Seaborn (33) were used to create the figures for data visualization.

Machine Learning Models and Evaluations

Seven machine learning models were developed to predict SBA utilization: logistic regression, decision tree, random forest, k-nearest neighbors (KNN), linear support vector machine (SVM), XGBoost, and Naive Bayes (31). Hyperparameter tuning for each model was performed using GridSearchCV with 5-fold stratified cross-validation on the SMOTE-balanced training dataset, optimizing for the area under the receiver operating characteristic curve (AUC).

The best-performing hyperparameters identified through GridSearchCV were then used to train the final models. Model performance was evaluated on the independent test dataset using accuracy, sensitivity, specificity, F1-score, and AUC.

Receiver operating characteristic (ROC) curves were generated to compare discrimination performance across models.

To assess model robustness and generalizability, additional stratified k-fold cross-validation (10, 15, and 20-fold) was conducted using the optimal hyperparameters (24). To prevent data leakage, SMOTE was incorporated within a pipeline during cross-validation, ensuring resampling was applied only to training folds.

This approach ensured stable performance across different data partitions and reduced the risk of overfitting.

Logistic regression and interpretation of features

To enhance interpretability, a binary logistic regression model was fitted using the final selected features to estimate adjusted associations between predictors and SBA utilization. Regression coefficients were exponentiated and reported as odds ratios (ORs) to quantify effect sizes (26).

Additionally, SHAP (SHapley Additive exPlanations) was applied to the XGBoost model to assess feature importance and model interpretability. SHAP values were used to quantify feature contributions, rank predictor importance, and explore interaction effects, providing insight into complex relationships captured by machine learning models.

The full methodological workflow diagram of the proposed hybrid analytical framework is shown in Fig. 1 .

Ethical Consideration

This study utilized publicly available, de-identified secondary data from the Bangladesh Demographic and Health Survey (BDHS) 2022. The BDHS survey protocols were reviewed and approved by the Institutional Review Boards (IRBs) of the National Institute of Population Research and Training (NIPORT), ICF, and icddr,b, and informed consent was obtained from all participants prior to data collection. As this study involved secondary analysis of anonymized data with no identifiable personal information, no additional ethical approval was required. The use of the dataset complied with the guidelines of the DHS Program.

Results

A total of 72.9% of women utilized skilled birth attendants (SBAs) in the study population. Table 2 presents the predictive performance of seven machine learning models for skilled birth attendant (SBA) utilization.

Table 2
Performance comparison of machine learning models for SBA prediction

Model	Accuracy	Sensitivity	Specificity	F1 Score	AUC
Logistic Regression	0.71	0.72	0.71	0.78	0.77
Decision Tree	0.72	0.84	0.42	0.81	0.66
Random Forest	0.73	0.88	0.37	0.83	0.74
KNN	0.63	0.65	0.62	0.72	0.67
Linear SVM	0.71	0.72	0.71	0.78	0.77
XGBoost	0.75	0.91	0.33	0.84	0.75
Naive Bayes	0.70	0.72	0.69	0.77	0.75

XGBoost achieved the highest accuracy (0.75), sensitivity (0.91), and F1-score (0.84), indicating superior performance in identifying SBA users, although with relatively low specificity (0.33). Random Forest demonstrated a similar pattern, with high sensitivity (0.88) and F1-score (0.83), but limited specificity (0.37). Logistic Regression and linear support vector machine (SVM) exhibited balanced performance across all evaluation metrics, each achieving an accuracy of 0.71, sensitivity of 0.72, specificity of 0.71, and the highest AUC (0.77). In contrast, the Decision Tree model showed high sensitivity (0.84) but lower specificity (0.42) and AUC (0.66), indicating weaker overall discriminative ability. K-Nearest Neighbors (KNN) demonstrated the lowest performance (accuracy = 0.63, AUC = 0.67), whereas Naive Bayes showed moderate and stable results across all metrics. Although logistic regression and SVM achieved the highest AUC (0.77), XGBoost was considered the best-performing model due to its superior accuracy, sensitivity, and F1-score.

Figure 2 illustrates the receiver operating characteristic (ROC) curves of the models.

Logistic Regression and Linear SVM achieved the highest discriminative ability (AUC = 0.77), followed by XGBoost and Naive Bayes (AUC = 0.75), and Random Forest (AUC = 0.74). In contrast, KNN (AUC = 0.67) and Decision Tree (AUC = 0.66) demonstrated comparatively lower predictive performance. The ROC curves further indicate that Logistic Regression and SVM maintain a more consistent balance between sensitivity and specificity across different thresholds, whereas other models exhibit relatively weaker discrimination.

Table 3 presents the cross-validation results using 10, 15, and 20-fold validation.

Table 3
Cross-validation performance of machine learning models for SBA prediction

Model	Accuracy	Sensitivity	Specificity	F1 Score	AUC
10-Fold					
Logistic Regression	0.70	0.69	0.71	0.77	0.77
Decision Tree	0.71	0.79	0.49	0.80	0.71
Random Forest	0.75	0.86	0.46	0.83	0.76
KNN	0.65	0.65	0.66	0.73	0.70
Linear SVM	0.70	0.69	0.71	0.77	0.77
XGBoost	0.75	0.89	0.38	0.84	0.77
Naive Bayes	0.69	0.71	0.63	0.77	0.74
15-Fold					
Logistic Regression	0.70	0.69	0.71	0.77	0.77
Decision Tree	0.71	0.80	0.46	0.80	0.70
Random Forest	0.75	0.86	0.46	0.83	0.76
KNN	0.65	0.65	0.66	0.73	0.70
Linear SVM	0.70	0.69	0.71	0.77	0.77
XGBoost	0.75	0.89	0.37	0.84	0.77
Naive Bayes	0.69	0.71	0.64	0.77	0.74
20-Fold					
Logistic Regression	0.70	0.70	0.72	0.77	0.77
Decision Tree	0.70	0.78	0.47	0.79	0.70
Random Forest	0.75	0.86	0.46	0.83	0.76
KNN	0.65	0.65	0.65	0.73	0.70
Linear SVM	0.70	0.69	0.72	0.77	0.77
XGBoost	0.75	0.89	0.38	0.84	0.77
Naive Bayes	0.69	0.71	0.65	0.77	0.74

The performance of all models remained highly consistent across different folds, indicating strong robustness and model stability. XGBoost and Random Forest consistently demonstrated superior performance, with high accuracy (≈ 0.75), sensitivity (≈ 0.86 – 0.89), and F1-score (≈ 0.83 – 0.84), although both exhibited relatively low specificity (≈ 0.37 – 0.46). Logistic Regression and Linear SVM showed stable and balanced performance across all folds, maintaining consistent accuracy (≈ 0.70), specificity (≈ 0.71 – 0.72), and the highest AUC values (≈ 0.77).

Decision Tree showed moderate performance with relatively high sensitivity but lower specificity and AUC, while KNN consistently yielded the lowest performance. Naive Bayes demonstrated moderate and stable results across all validation settings. The minimal variation across folds suggests that the models are not overfitting and that the findings are generalizable to unseen data.

Figure 3 presents the SHAP (SHapley Additive exPlanations) summary plot for the XGBoost model, illustrating the contribution of individual features to SBA utilization predictions. Features are ranked according to their impact on the model output.

Antenatal care (ANC)-related variables emerged as the most influential predictors. Having at least one or two ANC visits and early initiation of ANC (within the first trimester) positively associated with SBA utilization. Socioeconomic factors also played a significant role, with wealth status and media exposure positively associated with SBA use, while poverty reduced the likelihood. Demographic and

reproductive characteristics further influenced predictions. Higher parity (≥ 3 children) and receiving ANC from unskilled providers were associated with lower SBA utilization. Regional variation was also observed, indicating geographic disparities in service use.

Table 4 presents the logistic regression results identifying key determinants of SBA utilization.

Table 4
Logistic regression results for determinants of SBA utilization

Variable	Category	Beta	Odds Ratio
Division	Khulna	1.41	4.11
Antenatal Care Visits	Once	-1.09	0.34
Religion	Non-Muslim	1.05	2.86
Education Level	Primary	-0.89	0.41
Antenatal Care Provider	Unskilled Provider	-0.82	0.44
Age Group	45–49	-0.79	0.46
Antenatal Care Visits	Two Times	-0.72	0.49
Education Level	Secondary	-0.71	0.49
Number of Living Children	≥ 3 Children	-0.58	0.56
Education Level	No Education	-0.54	0.58
Division	Rajshahi	0.49	1.64
Wealth Index	Poor	-0.41	0.66
Media Exposure	Yes	0.33	1.39
Wealth Index	Rich	0.32	1.38
Division	Mymensingh	-0.31	0.74
Husband Occupation	Service Holder	0.29	1.34
Division	Dhaka	0.29	1.34
Husband Occupation	Unemployed	0.29	1.34
Number of Living Children	2 Children	-0.29	0.75
Age at First Birth	≤ 20 Years	-0.27	0.76

Several factors were positively associated with SBA use, including residence in Khulna division (OR = 4.11), Rajshahi division (OR = 1.64), non-Muslim religion (OR = 2.86), media exposure (OR = 1.39), and higher wealth status (OR = 1.38). Conversely, reduced likelihood of SBA utilization was observed among women with fewer ANC visits, particularly those with only one (OR = 0.34) or two visits (OR = 0.49), as well as those receiving care from unskilled providers (OR = 0.44). Lower educational attainment, higher parity (≥ 3 children), older maternal age (45–49 years), and early age at first birth (≤ 20 years) were also negatively associated with SBA use. Additionally, women from poorer households were less likely to utilize SBA services (OR = 0.66).

These results indicate that geographic location, socioeconomic status, media exposure, and adequate antenatal care are associated with SBA utilization, while lower education, higher fertility, and limited access to skilled care reduce the likelihood of service use.

Overall, both statistical and machine learning findings consistently highlight the importance of healthcare access and socioeconomic inequality in shaping SBA utilization in Bangladesh.

Discussion

This paper combines conventional statistical modeling and ensemble machine learning to determine the key determinants of skilled birth attendant (SBA) use among 4,646 women in Bangladesh, based on the Bangladesh Demographic and Health Survey (BDHS) 2022. Logistic

regression analysis revealed strong associations, with women residing in Khulna division (OR = 4.11) and non-Muslim women (OR = 2.86) significantly more likely to utilize SBAs. On the other hand, one visit to antenatal care (ANC) was associated with a 66% decrease in SBA utilization (OR = 0.34). These findings were supported by machine learning models, with XGBoost demonstrating the best overall predictive performance in terms of accuracy, sensitivity, and F1-score, indicating its effectiveness in identifying SBA users, while logistic regression and SVM achieved the highest discriminative ability based on AUC (0.77). Given the public health importance of correctly identifying SBA users, sensitivity and F1-score were prioritized when determining the best-performing model.

Notably, ensemble models such as XGBoost and Random Forest exhibited high sensitivity but relatively low specificity, indicating a greater ability to correctly identify SBA users than non-users. This imbalance suggests that while these models are effective in identifying women likely to utilize SBAs, they are less effective in distinguishing non-users, reflecting a prioritization of sensitivity over specificity, which is often desirable in public health screening contexts where missing true cases may have serious consequences.

Such trade-offs are common in predictive modeling of public health outcomes and should be considered when applying these models for policy targeting and intervention design (34, 35). The consistency of model performance across 10, 15, and 20-fold cross-validation further supports the robustness and generalizability of the findings. The minimal variation in performance metrics across folds indicates that the models are stable, not overfitting, and capable of generalizing to unseen data, thereby strengthening confidence in their predictive capability (24).

Our results are consistent with the current evidence on Bangladesh and other low and middle-income country (LMIC) settings, highlighting the role of socioeconomic and healthcare gradients in SBA uptake. Household wealth also became one of the most significant determinants, with wealthier households having higher chances of SBA utilization (OR = 1.38), similar to the works of Bhowmik et al. (13) and Afroja et al. (36). Mother education also had a positive effect, and the odds ratios of lower levels of education, including primary (OR = 0.41) and secondary (OR = 0.49), were lower than those of higher levels of education (12, 11). These findings are consistent with international evidence, e.g., by Dickson et al. (8) and Walker et al. (37), who reported educational empowerment as a predictor of SBA in sub-Saharan African nations. Media and internet exposure, combined with wealth and education, increased the propensity toward SBA utilization, highlighting the multidimensional nature of access to maternal healthcare.

The timing and frequency of antenatal care (ANC) visits are significant and modifiable predictors of skilled birth attendant (SBA) use. Women who initiate ANC in the first trimester and complete the recommended number of contacts of ANC visits are more likely to use SBAs, which is consistent with global evidence and country analyses (38, 39). In Bangladesh, multiple studies also show that receiving ANC (and receiving more ANC) is associated with greater SBA use (13, 40). The evidence suggests a dose-response pattern, where SBA use increases with additional ANC contacts, aligning with the shift from ≥ 4 visits to the WHO's 8-contact model (38, 13). In addition, women who received ANC from unskilled providers were significantly less likely to utilize SBAs, indicating gaps in the continuum and quality of maternal healthcare services. The SHAP-based feature importance analysis further reinforced these findings by identifying ANC-related variables as the most influential predictors of SBA utilization, highlighting the importance of both access to and quality of maternal health services.

Geographic and urban-rural disparities remained evident, with greater utilization of SBAs in urban regions and in divisions such as Khulna and Rajshahi than in Mymensingh and Sylhet (6). In addition, women from Rajshahi division were more likely to utilize SBAs (OR = 1.64), whereas those from Mymensingh showed lower utilization, indicating persistent regional disparities in maternal healthcare access. As demonstrated by Fredriksson et al. (34), machine learning models are capable of quantifying inequalities and surfacing interactions among socioeconomic variables. Moreover, there is mounting support that media exposure and internet access are becoming determinants of maternal care utilization in low- and middle-income environments and indicators of the evolving position of digital health (41). These factors enable women to access timely health information, overcome cultural barriers, and make informed healthcare choices.

Religion had location-specific impacts: non-Muslim women in Bangladesh had a higher likelihood of SBA utilization, and similar patterns have been observed in other contexts, such as Guinea (42), highlighting the impact of cultural and social factors. These predictors also describe resource-constrained environments in which these mechanisms are reinforcing: wealth facilitates access, education may promote decision-making and health literacy, ANC visits create trust in formal facilities, and media exposure improves access to information (43, 16, 13).

Demographic and reproductive factors also played a significant role. Higher parity (≥ 3 children) was associated with lower SBA utilization, possibly due to increased confidence in home delivery or reduced perceived need for skilled care. Similarly, older maternal age (45–49 years) and early age at first birth (≤ 20 years) were negatively associated with SBA use, reflecting entrenched behavioral patterns and generational differences in maternal health practices (16).

This study makes an important methodological contribution, as it demonstrates how interpretable logistic regression can be used together with high-performing ensemble machine learning models such as random forest and XGBoost to capture complex and nonlinear relationships and interactions (34, 44). This hybrid framework improves both predictive accuracy and interpretability, thereby enhancing its applicability for evidence-based public health decision-making.

Importantly, these findings are directly aligned with the Sustainable Development Goal (SDG) target 3.1.2, which emphasizes increasing the proportion of births attended by skilled health personnel. The identification of key determinants such as ANC utilization, maternal education, and socioeconomic status provides actionable insights for accelerating progress toward this global target in Bangladesh (6).

Strengths, Limitations, and Policy Implications

This study has several strengths. It utilized a nationally representative BDHS 2022 dataset, enhancing the generalizability of findings across Bangladesh. A robust hybrid feature selection approach combining statistical and machine learning techniques was applied to identify key predictors. The use of SMOTE addressed class imbalance, while multiple k-fold cross-validation ensured model stability and reduced overfitting. Additionally, the integration of SHAP analysis improved model interpretability, bridging the gap between predictive performance and explainability.

However, the study has some limitations. First, the cross-sectional design limits causal inference. Second, the reliance on self-reported data may introduce recall bias. Third, important contextual variables such as healthcare quality, accessibility, and cultural norms were not included in the dataset. Finally, although machine learning models improved predictive performance, their real-world implementation in healthcare systems requires further validation.

The findings of this study have important policy implications for improving maternal healthcare utilization in Bangladesh. First, strengthening antenatal care (ANC) services should be a priority, particularly promoting early initiation and ensuring adequate number of visits. Second, targeted interventions should focus on disadvantaged populations, including women from poorer households and those with lower educational attainment. Third, region-specific strategies are required to reduce disparities in divisions such as Mymensingh and Sylhet, where SBA utilization remains comparatively low. Fourth, expanding media-based and digital health interventions can enhance awareness and improve healthcare-seeking behavior. Finally, the integration of machine learning approaches into national health systems may facilitate data-driven identification of high-risk groups and support efficient allocation of healthcare resources, thereby accelerating progress toward achieving SDG 3.1.2.

Conclusions

The use of skilled birth attendants (SBAs) in Bangladesh is associated with socioeconomic status, maternal education, antenatal care (ANC) utilization, media exposure, and geographic disparities. The integrated application of machine learning and statistical approaches showed that XGBoost achieved the best overall predictive performance in terms of accuracy, sensitivity, and F1-score, indicating its effectiveness in identifying SBA users, while logistic regression and support vector machine (SVM) demonstrated the highest discriminative ability based on AUC and provided interpretable, policy-relevant insights. The findings further highlight the critical role of adequate and timely ANC, including early initiation and sufficient visits, as well as the importance of service quality in promoting SBA utilization. Targeted interventions should therefore focus on expanding ANC coverage, improving quality of care, strengthening media-based awareness, and reducing regional and socioeconomic inequalities. Future research should incorporate longitudinal and spatial modeling approaches to enhance causal inference and policy relevance. Overall, the analytical framework developed in this study provides a robust, evidence-based approach to support data-driven policymaking and the design of targeted public health interventions in Bangladesh.

Abbreviations

ANC: Antenatal care

AUC: Area under the receiver operating characteristic curve

BDHS: Bangladesh Demographic and Health Survey

CSBA: Community Skilled Birth Attendant

DHS: Demographic and Health Surveys (DHS) Program

EA: Enumeration Areas (EAs)

FWV: Family Welfare Visitors (FWVs)

KNN: K-nearest neighbors

LMIC: Low- and middle-income country

ML: Machine learning

NIPORT: National Institute of Population Research and Training

OR: Odds ratio

ROC: Receiver operating characteristic

SACMO: Sub-assistant Community Medical Officer (SACMOs)

SBA: Skilled Birth Attendant (SBAs)

SDG: Sustainable Development Goal

SVM: Support vector machine

SMOTE: Synthetic Minority Oversampling Technique

USAID: United States Agency for International Development

WHO: World Health Organization

XGBoost: Extreme Gradient Boosting

Declarations

Ethics approval and consent to participate

This study used publicly available, de-identified secondary data from the Bangladesh Demographic and Health Survey (BDHS) 2022. The BDHS protocols were reviewed and approved by the Institutional Review Boards (IRBs) of the National Institute of Population Research and Training (NIPORT), ICF, and icddr,b. Informed consent was obtained from all participants prior to data collection. Permission to use the dataset was obtained from the DHS Program. No additional ethical approval was required for this secondary data analysis.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analyzed during the current study are publicly available from the DHS Program repository (<https://dhsprogram.com>) upon reasonable request and registration.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no specific funding for this work.

Author contributions

Conceptualization: Md. Rashidul Islam Rifat, Md. Deluar Jahan Moloy.

Data curation: Md. Rashidul Islam Rifat, Arzo Ahmed, **Most. Farjana Yesmin.**

Formal analysis: Md. Rashidul Islam Rifat.

Methodology: Md. Rashidul Islam Rifat, Md. Deluar Jahan Moloy.

Software: Md. Rashidul Islam Rifat.

Validation: Md. Deluar Jahan Moloy, Md. Mizanur Rahman.

Visualization: Md. Rashidul Islam Rifat, Tasnia Rahman Shefa, Md. Onthor.

Supervision: Md. Deluar Jahan Moloy.

Writing – original draft: Md. Rashidul Islam Rifat.

Writing – review & editing: Md. Deluar Jahan Moloy, Arzo Ahmed, Md. Mizanur Rahman, Most. Farjana Yesmin, Md. Onthor, Tasnia Rahman Shefa.

Acknowledgments

We thank the DHS Program for providing access to the BDHS 2022 dataset. The authors received no specific funding for this work.

References

1. World Health Organization. Maternal mortality [Fact sheet] [Internet]. 2024 [cited 2025 Nov 9]. Available from: <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>
2. World Bank. Maternal mortality ratio (modeled estimate, per 100,000 live births) – Bangladesh (indicator SH.STA.MMRT) [Dataset] [Internet]. 2023. Available from: <https://databank.worldbank.org/source/world-development-indicators>
3. WHO. Definition of skilled health personnel providing care during childbirth [Internet]. Geneva, Switzerland: World Health Organization; 2018 [cited 2025 Nov 9]. Available from: <https://www.who.int/publications/i/item/WHO-RHR-18.14>
4. Bhutta ZA, Das JK, Bahl R, Lawn JE, Salam RA, Paul VK, et al. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *Lancet*. 2014;384(9940):347–70. 10.1016/S0140-6736(14)60792-3.
5. Wagstaff A, Claeson M. The Millennium Development Goals for Health: Rising to the Challenges [Internet]. Washington, DC: World Bank; 2004 [cited 2025 Nov 9]. Available from: <https://openknowledge.worldbank.org/handle/10986/14954>
6. Chowdhury SSA, Kundu S, Sharif AB. Socioeconomic and geographical inequalities in using skilled birth attendants during delivery in Bangladesh over two decades. *BMC Pregnancy Childbirth*. 2023;23(1):430. 10.1186/s12884-023-05754-8.
7. Bobo FT, Asante A, Woldie M, Dawson A, Hayen A. Spatial patterns and inequalities in skilled birth attendance and caesarean delivery in sub-Saharan Africa. *BMJ Glob Health*. 2021;6(10):e007074. 10.1136/bmjgh-2021-007074.
8. Dickson KS, Adde KS, Ameyaw EK. Women empowerment and skilled birth attendance in sub-Saharan Africa: A multi-country analysis. Kabir E, editor. *PLoS ONE*. 2021;16(7):e0254281. 10.1371/journal.pone.0254281
9. Bangladesh Bureau of Statistics (BBS); United Nations Population Fund (UNFPA). Bangladesh Sample Vital Registration System (SVRS) 2020 [Official statistical report] [Internet]. Dhaka, Bangladesh. 2020 [cited 2025 Nov 9]. Available from: https://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd/page/bdcb9aee_8bbd_45a2_a8f7_47d72c8a8c74/SVRS_2020.pdf
10. Ministry of Health and Family Welfare (MoHFW); ICF NI of PR and T (NIPORT). Bangladesh Demographic and Health Survey 2022: Key Indicators Report [Internet]. Dhaka, Bangladesh, and Rockville, Maryland, USA: National Institute of Population Research and Training (NIPORT); Ministry of Health and Family Welfare (MoHFW); ICF; 2024 [cited 2025 Nov 9]. Available from: <https://dhsprogram.com/publications/publication-BR141-Indicator-Profiles.cfm>
11. Nahar MT, Anik SMFI, Islam MA, Islam SMS. Individual and community-level factors associated with skilled birth attendants during delivery in Bangladesh: A multilevel analysis of demographic and health surveys. Kabir E, editor. *PLoS ONE*. 2022;17(6):e0267660.

12. Ema NJ, Eva MK, Mamun ASMA, Rafi MA, Khabir A, Kundu RN et al. Two level logistic regression analysis of factors influencing skilled birth attendant during delivery among Bangladeshi mothers: A nationally representative sample. *Wekere FCC, editor. PLoS ONE.* 2023;18(9):e0291790. 10.1371/journal.pone.0291790
13. Bhowmik J, Biswas RK, Woldegiorgis M. Antenatal care and skilled birth attendance in Bangladesh are influenced by female education and family affordability: BDHS 2014. *Public Health.* 2019;170:113–21. 10.1016/j.puhe.2019.02.027.
14. Rahman MO, Rauf MA, Ulfa Y, Siddiqi MNA, Islam MR, Inaoka K et al. Association of quality antenatal care and completion of eight or more antenatal care visits with skilled delivery care utilization among pregnant women in Bangladesh: A nationwide population-based study. *Fufa DA, editor. PLoS One.* 2025;20(4):e0322725. 10.1371/journal.pone.0322725
15. Setu SP, Islam MA. Individual and Community-Level Determinants of Institutional Delivery Services among Women in Bangladesh: A Cross-Sectional Study. *Int J Clin Pract.* 2022;2022(1):3340578. 10.1155/2022/3340578.
16. Rahman MA, Kundu S, Rashid HO, Shanto HH, Rahman MM, Khan B, et al. Socioeconomic inequalities in utilizing facility delivery in Bangladesh: A decomposition analysis using nationwide 2017–2018 demographic and health survey data. *PLoS ONE.* 2022;17(11):e0278093. 10.1371/journal.pone.0278093.
17. Mennickent D, Rodríguez A, Opazo MC, Riedel CA, Castro E, Eriz-Salinas A, et al. Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications. *Front Endocrinol.* 2023;14. 10.3389/fendo.2023.1130139.
18. Islam MN, Mustafina SN, Mahmud T, Khan NI. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. *BMC Pregnancy Childbirth.* 2022;22(1):348. 10.1186/s12884-022-04594-2.
19. Sultana N, Afia Z, Sifat IK, Zoha S, Jisa TA, Kibria MK. Machine learning based prediction of low birth weight and its associated risk factors: Insights from the Bangladesh Demographic and Health Survey 2022. *Babu GR, editor. PLOS Glob Public Health.* 2025;5(9):e0005187. 10.1371/journal.pgph.0005187
20. Bhowmik J, Biswas RK, Ananna N. Women's education and coverage of skilled birth attendance: An assessment of Sustainable Development Goal 3.1 in the South and Southeast Asian Region. *Wilunda C, editor. PLoS ONE.* 2020;15(4):e0231489. 10.1371/journal.pone.0231489
21. NIPORT ICF. Bangladesh Demographic and Health Survey 2022 [Demographic and Health Survey] [Internet]. National Institute of Population Research and Training (NIPORT) and ICF; 2023. Available from: <https://dhsprogram.com>
22. ICF. Demographic and Health Surveys (DHS) Program. DHS Program [Internet]. 2023. Available from: <https://dhsprogram.com>
23. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17(3):261–72. 10.1038/s41592-019-0686-2.
24. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Vol. 2. Montreal, Canada: Morgan Kaufmann; 1995. pp. 1137–43. Available from: <https://doi.org/10.5555/1643031.1643047>
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res.* 2002;16:321–57. 10.1613/jair.953.
26. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression [Internet]. 1st ed. Wiley; 2013 [cited 2025 Nov 25]. (Wiley Series in Probability and Statistics). Available from: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118548387> doi:10.1002/9781118548387.
27. O'brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual Quant.* 2007;41(5):673–90. 10.1007/s11135-006-9018-6.
28. McKinney W. Data Structures for Statistical Computing in Python. In. Austin, Texas; 2010 [cited 2025 Nov 25]. pp. 56–61. Available from: <https://doi.curvenote.com/10.25080/Majora-92bf1922-00a> doi:10.25080/Majora-92bf1922-00a.
29. Harris CR, Millman KJ, Van Der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature.* 2020;585(7825):357–62. 10.1038/s41586-020-2649-2.
30. Pedregosa F, Pedregosa F, Varoquaux G, Varoquaux G, Org N, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON;* 2011.
31. Chen T, Guestrin C, XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. San Francisco California USA: ACM; 2016 [cited 2025 Nov 25]. pp. 785–94. Available from: <https://dl.acm.org/doi/10.1145/2939672.2939785> doi:10.1145/2939672.2939785.
32. Hunter JD, Matplotlib. A 2D Graphics Environment. *Comput Sci Eng.* 2007;9(3):90–5. 10.1109/MCSE.2007.55.
33. Waskom M. seaborn: statistical data visualization. *JOSS.* 2021;6(60):3021. 10.21105/joss.03021.

34. Fredriksson A, Fulcher IR, Russell AL, Li T, Tsai YT, Seif SS, et al. Machine learning for maternal health: Predicting delivery location in a community health worker program in Zanzibar. *Front Digit Health*. 2022;4:855236. 10.3389/fdgth.2022.855236.
35. Kuhn M, Johnson K, New York. *Applied Predictive Modeling* [Internet]. New York, NY: Springer ; 2013 [cited 2025 Dec 1]. Available from: <http://link.springer.com/10.1007/978-1-4614-6849-3> doi:10.1007/978-1-4614-6849-3.
36. Afroja S, Muhammad Nasim AS, Khan MS, Kabir MA. Rural-Urban Determinants of Receiving Skilled Birth Attendants among Women in Bangladesh: Evidence from National Survey 2017-18. Parameshachari BD, editor. *International Journal of Clinical Practice*. 2022;2022(1):5426875. 10.1155/2022/5426875
37. Walker T, Woldegiorgis M, Bhowmik J. Utilisation of Skilled Birth Attendant in Low- and Middle-Income Countries: Trajectories and Key Sociodemographic Factors. *IJERPH*. 2021;18(20):10722. 10.3390/ijerph182010722.
38. Jiwani SS, Amouzou-Aguirre A, Carvajal L, Chou D, Keita Y, Moran AC, et al. Timing and number of antenatal care contacts in low and middle-income countries: Analysis in the Countdown to 2030 priority countries. *J Global Health*. 2020;10(1):010502. 10.7189/jogh.10.010502.
39. Woldeamanuel BT, Belachew TA. Timing of first antenatal care visits and number of items of antenatal care contents received and associated factors in Ethiopia: multilevel mixed effects analysis. *Reprod Health*. 2021;18(1):233. 10.1186/s12978-021-01275-9.
40. Kibria GMA, Ghosh S, Hossen S, Barsha RAA, Sharmeen A, Uddin SMI. Factors affecting deliveries attended by skilled birth attendants in Bangladesh. *matern health neonatol perinatol*. 2017;3(1):7. 10.1186/s40748-017-0046-0.
41. Mohamed H, Ismail A, Sutan R, Rahman RA, Juval K. A scoping review of digital technologies in antenatal care: recent progress and applications of digital technologies. *BMC Pregnancy Childbirth*. 2025;25(1):153. 10.1186/s12884-025-07209-8.
42. Ahinkorah BO, Seidu AA, 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 Childbirth*. 2021;21(1):2. 10.1186/s12884-020-03489-4.
43. Joseph G, Da Silva ICM, Wehrmeister FC, Barros AJD, Victora CG. Inequalities in the coverage of place of delivery and skilled birth attendance: analyses of cross-sectional surveys in 80 low and middle-income countries. *Reprod Health*. 2016;13(1):77. 10.1186/s12978-016-0192-2.
44. Taye EA, Woubet EY, Hailie GY, Arage FG, Zerihun TE, Zegeye AT, et al. Application of the random forest algorithm to predict skilled birth attendance and identify determinants among reproductive-age women in 27 Sub-Saharan African countries; machine learning analysis. *BMC Public Health*. 2025;25(1):901. 10.1186/s12889-025-22007-9.

Figures

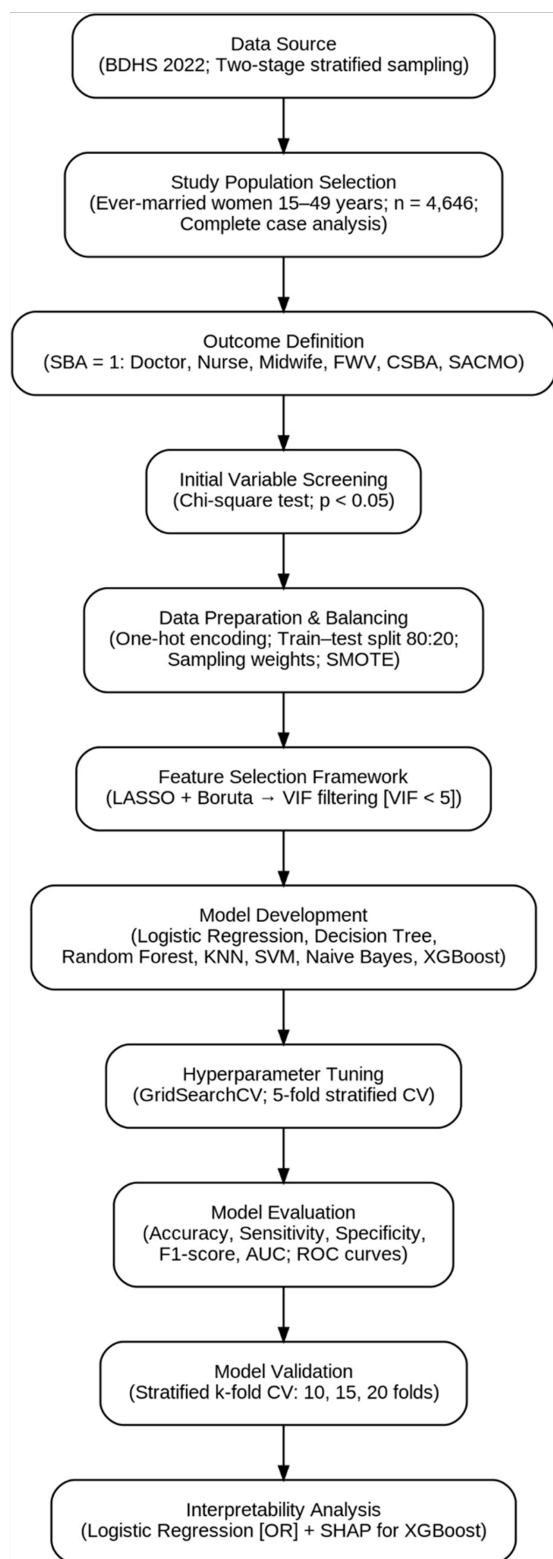


Figure 1

Hybrid analytical workflow for SBA utilization prediction

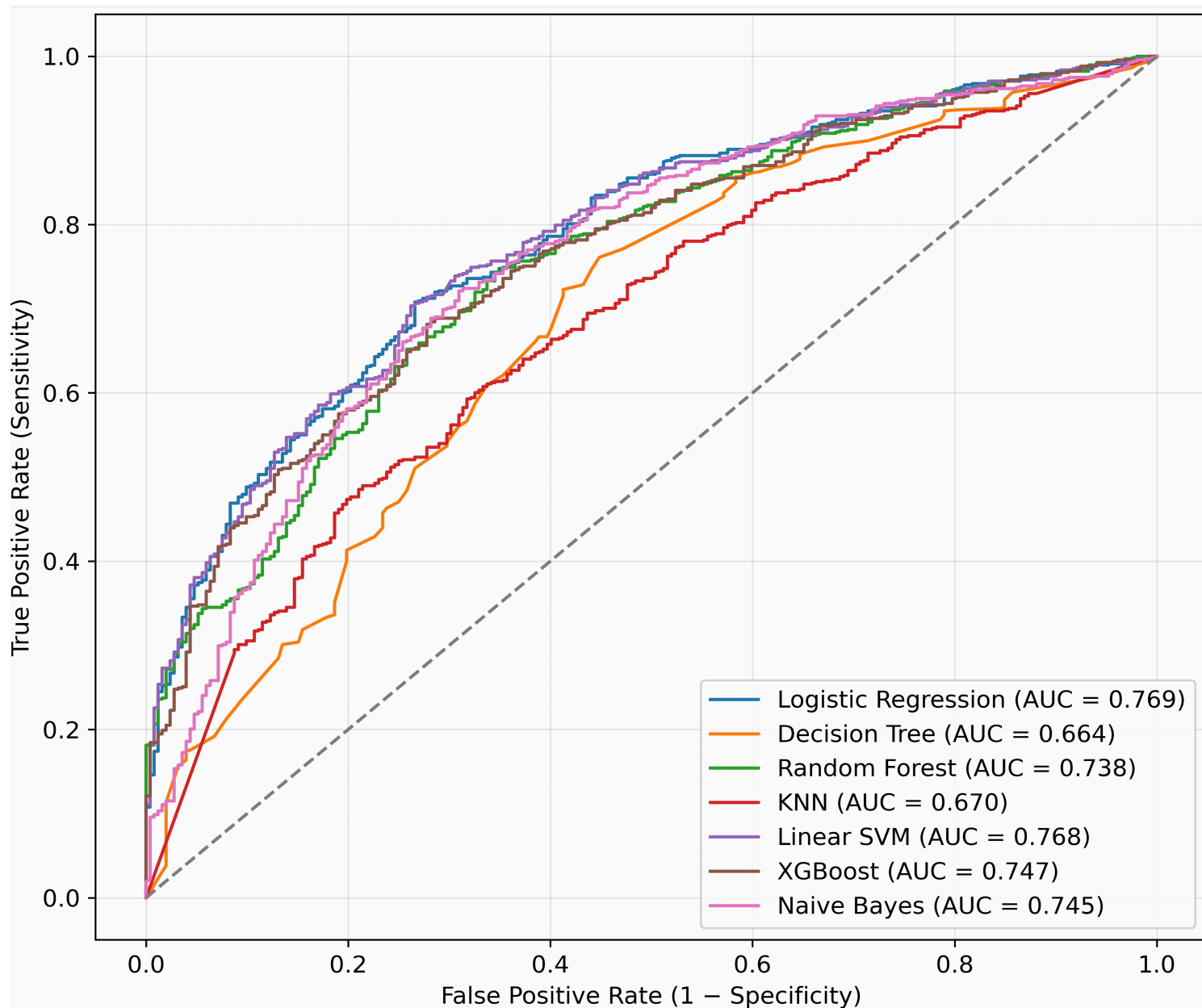


Figure 2

ROC curves of machine learning models for SBA prediction

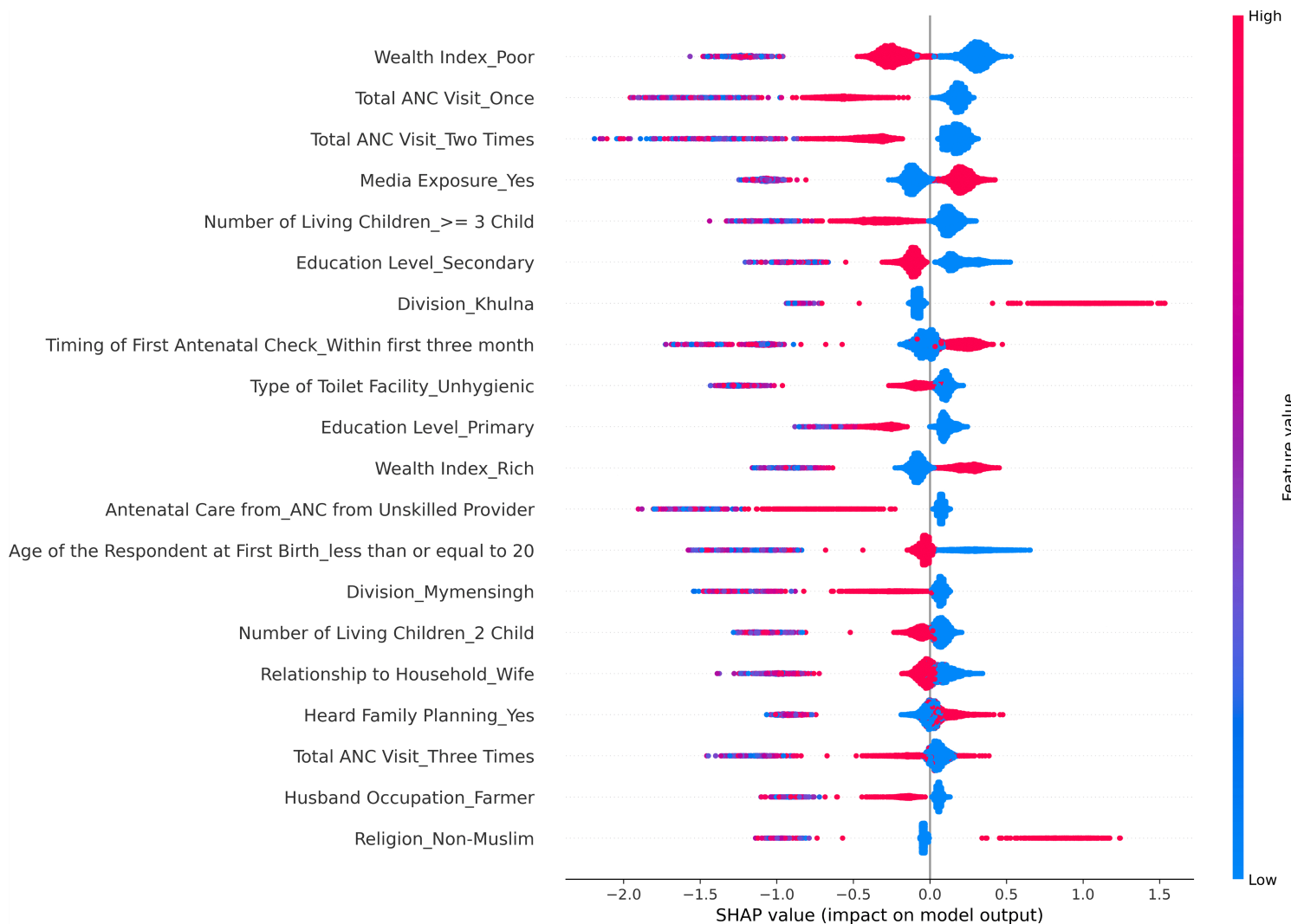


Figure 3

SHAP-based feature importance for SBA utilization prediction