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.

Identifying Key Determinants and Spatial Insights of Skilled Birth Assistance in Bangladesh: A Machine Learning Approach from BDHS 2022

Md Salek Miah

2021134066@student.sust.edu

Shahjalal University of Science & Technology (SUST)

Fhameda Faija Lamia

Sylhet Engineering College

Mst. Labonna Begum

Sylhet Engineering College

Article

Keywords: SBA, BDHS, Spatial, Machine Learning, SHAP.

Posted Date: August 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8029769/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

Objective

This study aims to identify the key determinants of skilled birth assistance in Bangladesh using machine learning and to examine spatial variation across regions.

Methodology

This study used data from the Bangladesh Demographic and Health Survey (BDHS) 2022, analyzing 4,871 ever-married women aged 15–49 years. Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbors, XGBoost, and Artificial Neural Network were trained to predict SBA. Model performance was assessed using accuracy, F1 score, Matthews's correlation coefficient (MCC), Cohen's Kappa, and area under the receiver operating characteristic curve (AUROC).

Results

XGBoost achieved the highest performance (accuracy = 0.84, F1 = 0.88, MCC = 0.63, AUROC = 0.91). The SHAP summary plot revealed that delivery mode (Caesarean section) and postnatal doctor contact exerted the strongest positive influences on SBA prediction, followed by media exposure, decision-making autonomy, ≥4 antenatal visits, husband education, internet use, and household wealth. Negative SHAP effects were observed for rural residence, multiple children, age at first sex (<18 years), and northeastern divisions. Dhaka and Chittagong show maximum SBA coverage, and Rangpur, Sylhet, Mymensingh, and Barisal divisions show the minimum.

Conclusion

Improving access to maternal healthcare and empowering women through education, digital inclusion, and socioeconomic progress to accelerate toward SDG 3.1.

Introduction

Maternal mortality and morbidity remain a major challenge to the public health in the world and especially in the low- and middle-class countries (LMICs), even though it has improved a lot in the last several decades [1,2]. The World Health Organization estimates that every year about 295,000 women die because of pregnancy and childbirth complications, with most of the deaths happening in LMICs [3]. Majority of maternal deaths can be observed to be avoidable through access to skilled care in time, effective medical intervention and proper health system support [4]. It is among the most important interventions to reduce the rate of maternal death, as each childbirth must be accompanied by a trained birth attendant (SBA) trained midwives, nurses, or physicians [5]. SBAs can detect complications, offer timely interventions and refer mothers to relevant health facilities thereby preventing postpartum hemorrhage and infections, obstetric labor, and other complications related to death [6]. Although SBA has

been acknowledged as a primary maternal health intervention worldwide, access inequality is observed, which is also indicative of healthcare infrastructure inequalities, socioeconomic factors, education, and cultural beliefs [7]. This is because factors that affect SBA utilization must be understood to ensure that the Sustainable Development Goal (SDG) of less than 70 maternal deaths per 100,000 live births in the world by the year 2030 is achieved [8]. In Bangladesh, the results of maternal health have been gradually increasing in the last 20 years through national health programmers and community-based interventions with the enhancement of antenatal care [9]. Bangladesh Maternal Mortality Survey 2020 recorded a maternal mortality ratio of 173 deaths per 100,000 live births, which is showing improvement, but still showing disparities with high-income nations [10]. Even with government campaigns to encourage institutional birth and coverage of SBA, a large percentage of women, particularly those in rural and socioeconomically disadvantaged societies, are still giving birth in their homes without the advice of experts [11]. The major barriers are the lack of access to healthcare centers, economic factors, low educational levels of the mothers, cultural beliefs, and sociocultural beliefs, which do not support the use of facilities to deliver [12]. In addition, SBA coverage in different parts of Bangladesh is geographically different with Dhaka and Khulna divisions registering higher coverage and Sylhet and Barishal registering lower coverage [13]. These geographical differences imply that both the demographic and geographical variables have to be studied so that the accessibility of skilled birth attendance is evenly distributed throughout the country [9]. The determinants of SBA in countries and Bangladesh have been investigated in a few researches mostly employing conventional statistical methods like logistic regression [14]. Mothers age, maternal education, household wealth, parity, antenatal care visits, place of residence, and distance to health facilities are some of the factors that are traditionally recognized to affect SBA use [15]. Indicatively, the Demographic and Health Survey (DHS) time and again reveal that women who are highly educated, are wealthy in their households, and those who live in urban environments have higher chances of receiving skilled care when giving birth [16]. On the same note, SBA is positively related to the use of antenatal care and exposure to health information campaigns [17]. But these conventional methods are constrained by factors such as the assumption of linearity and the failure to represent multifaceted interaction of a couple of risk factors [18]. In addition, most of the research is individual-level as well as used traditional regression model, lacking incorporation of spatial information, including variation in SBA prevalence at the division-level [19]. Accordingly, the comparative significance of geographic differences is under-investigated, which does not allow for making specific measures aimed at eliminating high-risk areas [20]. By using machine learning (ML) methods, it is possible to identify non-linear relationships, consider feature interactions, and incorporate spatial variables to get a more detailed picture of predictors of SBA [21,22]. Although the literature on maternal health in Bangladesh is on the increase, the gaps in approach, spatial understanding, and policy recommendations are still present [14,23–25]. The traditional regression models can fail to capture subtle relationships among demographic socioeconomic, cultural, and spatial variables [26]. Although other studies have determined some of the determinants, they seldom determine the relative contribution of these determinants in a manner that can be interpreted and implemented by policymakers[27]. Therefore, the study objectives are to: 1) identify

key determinants of SBA using machine learning approach 2) examine administrative division- wise spatial variation of SBA 3) recommend strategies to improve maternal healthcare and reduce maternal mortality in Bangladesh.

Methodology

Data Sources and Study design

The study involved the use of secondary cross-sectional data, which was the 9th Bangladesh Demographic and Health Survey (BDHS) 2022 that was a nationally representative cross-sectional survey. The Government of Bangladesh and the United States Agency for International Development (USAID) financed the survey and it was carried out by the Bangladeshi research firm Mitra and Associates in Dhaka. The data were collected in four stages with a duration of about four weeks each, between June 27 and December 12, 2022. A stratified sampling procedure was done in two phases. A total of 30 149 households were chosen out of 675 Enumeration Areas (EAs) 237 urban and 438 rural. Lastly, 30,078 of the ever-married women who were aged 15 to 49 years were interviewed successfully [16]. All methods were performed in accordance with the relevant guidelines and regulations, and the study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

Sample Size

To analyses this, we limited the sample to women with at least one live birth within the five years before the survey with their last child still alive. All women whose key variables had missing data were excluded and this left an analytic sample of 4,871 women [**see in Supplementary Figure S1**].

Target variable

The target variable was the Skilled Birth Attendance (SBA) which was the one that was delivered with the help of a trained health worker such as a doctor, nurse, midwife, or other qualified health workers. The delivery in the presence of the traditional birth attendants, family, neighbors, or none was regarded as unskilled. The last variable was a coded nominal variable (1 skilled birth attendant and 0 unskilled birth attendant) [16].

Data Preprocessing

Predictors with 40% missing values considered this study imputed using Multiple Imputation by Chained Equations (MICE) with predictive mean matching when using numeric variables (instead of logistic regression), logistic regression when using binary variables (instead of polytomous regression) [28]. Class imbalance was corrected using Synthetic Minority Over-sampling Technique (SMOTE) in the train dataset [29]. In an effort to solve the issue of class imbalance, the data were randomly divided into training and

test sets, 80: 20, after the imputation to ensure that 30.36 % of the deliveries had skilled personnel in them; and 69.64% had unskilled personnel in them.

Feature selection

Among 32 predictors the Boruta algorithm was used to select 23 important predictors that were used as features. The confirmed variables were: delivery_mode, postnatal care (pnc), wealth category (wealth_cat), antenatal care visits (ANC_visit), husband's education (husb_edu), child category (child_cat), parity, maternal education (mat_edu), media exposure (media), place of residence (residence), internet use (internet_use), household assets (hh_assets), administrative division (hh_division), health insurance (health_ins), household religion (hh_religion), toilet facility (toilet), drinking water source (water), fertility preference (fert_pref), maternal age at first birth (mat_age1st), household materials (hh_materials), age at first sexual intercourse (age_first_sex), decision-making (deci_make), and maternal age (mat_age) [**Figure 2**]. Multicollinearity was minimized by eliminating highly correlated variables (r > 0.90) and adjusted in model.

Model Development

Seven classifiers were trained to make the predictions of SBA: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), k-Nearest Neighbors (KNN), XGBoost (XGB), and Artificial Neural Networks (ANN). Decision Trees arrange the data using hierarchical recursive splitting whereas here in the random forest, they enhance prediction by combining more than one decision tree[30]. SVM determines the most optimal hyperplane to classify the data, LR calculates the probability of binary results with interpretable coefficients, KNN classifies instances on the basis of closest neighbors, XGBoost relies on gradient boosting and regularization to make accurate predictions and ANN identifies complicated non-linear tendencies with interconnected layers. SHAP (SHapley Additive exPlainations) analysis was used to determine the important determinants of SBA[31–33] . STATA 17 (StataCorp, College Station, TX, USA) was used to handle data and machine learning algorithms were executed in R Studio 4.5.1 (R Foundation, Vienna, Austria). This study-maintained STROBE guidelines.

Data Partitioning and Hyperparameter Tuning

Hyperparameters were optimized based on grid search for each model. Random Forest models changed the sample size of variables sampled at each split (mtry = 2, 4, 6, 8), Decision Trees changed the complexity parameter (cp = 0.001 -0.051), KNN changed the number of neighbors (k = 315), SVM changed the RBF kernel (sigma = 0.01), and cost parameter (C = 0.25-2), XGBoost changed learning rate (eta = 0.01, 0.1, 0.3), tree depth (Logistic regression did not necessitate any tuning of hyper-parameters .

Model Evaluation

Accuracy, precision, sensitivity (recall), specificity, F1-score, Cohen is Kappa and the use of the AUROC were used to determine model performance. Accuracy is the total percentage of correct results but can be not true when there is imbalance in the data [34]. Sensitivity is a measure of the proportion of true positives which have been correctly identified whereas the proportion of true predictions which are true positives is the measure of precision [35]. F1-score is a harmonic average of both precision and recall which is used in skewed datasets to give a balanced measure [36]. Kappa is a metric by Cohen that is used to measure agreement between observed and predicted classifications, correction of chance, and the overall discriminative ability is analyzed by AUROC [37]. When data is skewed, MCC is more preferable since it provides a fair image of the performance of a model on both minority and majority classes and does not provide the false impression as other measures of accuracy or precision do [30].

Statistical Analysis

All the predictors were calculated as descriptive statistics and the relationship of univariate with SBA was tested with chi-square tests concerning categorical variables. The preprocessing of the data, feature selection, and model training were based on the best practice of predictive modeling in healthcare outcomes that are imbalanced.

Results

The machine learning models used to predict skilled birth attendance, their main tuning parameters, and purpose. Each model was adjusted to improve accuracy, control complexity, and prevent overfitting [**Table 1**]. Skilled attendance was more prevalent in women between the age of 20 and 34, women with late age in the first birth (≥ 20 years) and those who entered sexual intercourse at the age of 18 and above. Maternal, husband education levels were strongly correlated with SBA and the women of non-Muslim households were found to have more coverage of SBA as compared to their Muslim counterparts. There was a significant relationship between SBA and urban residence, higher wealth (higher wealth than poor), fewer family members (5 or less), health insurance, household decision-making processes, household assets, and better household materials in terms of household and socioeconomic factors. Women living in cities, in richer families, having freedom in decision making, and having access to household resources had more chances of getting skilled birth care. The obstetric and reproductive factors, such as parity, ≥4 antenatal care (ANC) visits, mode of delivery, postnatal care (PNC), sexually active status, fertility preference, pregnancy loss, preterm birth, and current pregnancy status were strongly correlated with SBA. Women who had had at least 4 visits to ANC, women who delivered by caesarian, women who received PNC, and women who had been sexually active or wanted more children were better covered with SBA. Media exposure and empowerment variables such as any media exposure and internet use were also noted as significant and the women who were exposed to media or those who were internet users were more probable to receive skilled care. Lastly, geographic and context variables such as administrative division, type of toilet and source of drinking water were also important. Women in Dhaka, Khulna, and Barishal divisions, who had better sanitation and drinking

water, had higher chances to use the skilled birth services [**Table 2**]. Before applying SMOTE, the data providers (30.2%), the unskilled coverage (69.8%), showing an imbalance among the classes (Figure X). After using SMOTE, the two groups consisted of 50% of the sample, which allowed creating predictive models without bias and robustly **[Figure 1**]. XGBoost was showed the most overall performing algorithm with the highest accuracy (0.84), F1 score (0.88), Matthew's correlation coefficient (0.63), and the highest AUROC (0.91). In spite of the fact that Random Forest (accuracy = 0.82, AUROC = 0.91) and Logistic Regression (accuracy = 0.82, AUROC = 0.90) were also good at their work, XGBoost demonstrated a better balance between precision and recall, which shows that it was more stable during cross-class. Other models such as SVM, Decision Tree, KNN and Neural Network exhibited moderate predictive power with a minor unfavorable recall and calibration measures. On the whole, XGBoost is a chosen model because it has a high-accuracy level, high discriminative power, and a capability of describing nonlinear and complex relationships at the same time with low overfitting [**Table 3**].The XGBoost and the random forest models were found to be the best in terms of discriminative power on ROC curve analysis since they produced the best true positive rates at various thresholds [**Figure 3**]. The calibration plot of the XGBoost model showed that there was good similarity between predicted and observed success of the model, and the model is reliable to stratify risks among this population [**Figure 5**]. The most significant variable delivery mode (Caesarean) showed the highest positive SHAP influence, followed by PNC (Postnatal care), highlighting the strong effects of delivery and postnatal care on the predicted outcome. Positive SHAP contributions were also observed for media use, decision making, husband education as secondary and higher, ANC visits 4 times, internet use (yes), wealth category (Rich), age at first birth (<20 years), and fertility preference indicating that healthcare utilization, socioeconomic advantage, information access, and empowerment positively influence maternal health outcomes. In contrast, negative SHAP effects observed for residence(rural), administrative division (Mymensingh), number of child (multiple), and age at first sex suggesting that rural and regional disparities, higher parity, and early sexual debut reduce the predicted probability of favorable outcomes. A mild positive contribution was observed for division Chittagone **[Figure 4]**. Spatial The regional disparity is observed in the coverage of skilled providers in the whole Bangladesh, with the highest rates in Dhaka and Chittagone divisions (11.0-12.0%). Conversely, the coverage is lowest in Rangpur, Sylhet, Mymensingh and Barisal divisions (7.0-9.9%) [**Figure 6**].

Discussion

This paper utilized a combined machine learning and spatial analytical model to find out determinants of skilled birth attendance (SBA) in Bangladesh based on nationally representative data. The strong feature selection, interpretable machine learning modeling and geospatial mapping can provide an in-depth insight into maternal healthcare utilization patterns and disparities across the nation. It was found that there were significant regional and sociodemographic differences in SBA. Divisions with the highest coverage were Dhaka and Chittagong, as reflected on better infrastructure, literacy rates, and concentration of health care resources. On the other hand, such divisions as Rangpur, Sylhet, Mymensingh and Barisal showed constantly low coverage, which reflects inequity in access to qualified

health workers and facility-based delivery services in the regions. These results are in accordance with the previous analysis conducted by Bangladesh Demographic and Health Survey (BDHS), which revealed that the north and northeastern areas are under served because of poverty, transport access, and lack of health work force density [38]. XGBoost (Extreme Gradient Boosting) was identified as the most successful algorithm among all the tested algorithms. This is due to the fact that it achieves a higher performance because it has an ensemble learning algorithm which constructs several decision trees in a sequence which minimizes the classification error by gradient boosting and adds regularization to avoid overfitting [39]. In contrast to Logistic Regression, which presumes the linear relations between the predictors and the outcome, or Decision Trees, which may be erratic and susceptible to minor changes in the data, XGBoost is capable of capturing nonlinear and complex interactions between the sociodemographic, behavioral, and contextual determinants. Missing values, imbalance between classes (which is solved by SMOTE) and correlated features are other characteristics of the model that contribute to predictive reliability [40,41]. Notably, its stable accuracy-recall ratio means that it is effective in detecting both professional and non-professional deliveries and reduces bias due to the dominance of a class [42]. The better performance of XGBoost here is related to the results of the recent literature on the public health in low- and middle-income countries (LMICs), where the use of the ensemble-based models demonstrated high predictive accuracy of maternal and child health in low- and middle-income settings [43]. This has been supported by similar experiments in India, Kenya and Ethiopia which showed that boosting algorithms perform better than the traditional statistical models with large, heterogeneous data sets, thus confirming the selection of XGBoost as the best classifier to use in this research [27,44–47]. The Boruta algorithm of feature selection was used to eliminate all candidates except 23 statistically robust and relevant predictors that were included in the model. The domination of the mode of delivery as the most powerful predictor reflects the existence of a significant association between institutional delivery and skilled birth attendance. By definition, Caesarean section involves the use of skilled providers, and the fact that it has high predictive value proves that institutional deliveries are one of the key predictors of skilled attendance [18]. The spatial mapping showed that there were geographic clusters of poor SBA coverage in the north and northeast divisions. These tendencies reflect structural inequalities in that the rural districts experience poor transport networks, shortage of staff and delivery points [11]. Dhaka and Chittagong are geographical areas with concentrated healthcare networks that enjoy increased accessibility and availability of skilled work force [9,13]. The unending inequalities point to the urban rural divide and the effects of uneven health resource distributions [48]. The trends in the relationships between education and wealth and SBA are reminiscent of research in South Asia and sub-Saharan Africa, where the socioeconomic status is a strong predictor of institutional delivery [49,50]. The findings of Nepal, India and Ethiopia also confirm that better educated women, who have access to mass media and autonomy on which they make decisions, show much greater use of skilled providers [15,46,51–54]. Therefore, the present results can be combined with the existing knowledge on health system strengthening as the need to empower and bring about socioeconomic changes to enhance maternal health outcomes. The areas that lack geographic equity (especially in Mymensingh, Rangpur, and Sylhet) need targeted regional approaches, which could include better health facility access, transportation, and redistributed

skilled birth attendants. Increased women education, internet access, and access to financial services can directly help increase maternal health-seeking behavior. ANC, PNC, and delivery care should be incorporated into an uninterrupted service package to provide the holistic care of the mother. Midwifery and community-based outreach activities that are enhanced by mobile health technologies can be used to reach remote populations. Lastly, the surveillance systems that are machine learning-assisted may assist ministries to determine the regions that perform poorly and redistribute the resources in real time. The strengths of this research are that a large nationally representative dataset was used; feature selection (Boruta) was conducted effectively; machine learning (XGBoost) was used; and interpretation (SHAP values) was done clearly. Class imbalance bias was prevented by the use of SMOTE, whereas spatial visualization provided contextual understanding. Nonetheless, causal inference is restricted like other cross-sectional studies. It cannot be excluded that potential residual confounding and self-report measurement error may occur. Further, quality of care or cultural preferences, which are unmeasurable variables, could affect SBA utilization, but could not be obtained in the dataset.

Conclusions

This study reveals maternal healthcare-seeking behavior, educational attainment, household poverty, and regional inequalities are significant determinants of skilled birth attendance in Bangladesh. Of all the models evaluated in this study, the Extreme Gradient Boosting (XGBoost) model exhibited the most predictive power as well as the interpretability of complex nonlinear relationships of sociodemographic and spatial factors. Combining explainable machine learning with spatial analysis offers the opportunity for evidence-based, equitable resource targeting and spatially targeted maternal health interventions. Equity-increasing policies focused on improving maternal health systems, reducing rural-urban inequities, and social and economic empowerment of women, significantly in the high-burden regions of the country, will be crucial in pursuing Sustainable Development Goal 3.1.

Recommendations and Future Work

Skilled birth attendance can be improved by strengthening maternal health services, including antenatal, delivery, and postnatal care services in areas of low coverage, promotion of women's education, media exposure, and internet access to empower health-seeking behavior. The expansion of skilled workforce in underserved rural areas and machine learning-assisted monitoring is crucial to efficiently target interventions by addressing spatial disparities. Longitudinal trends, investigation into cultural factors of health service utilization, and assessment of quality of care are recommended for future studies. Application of predictive ML models with geospatial mapping to guide real-time policy and resource allocation for maternal and child health is also recommended.

Limitations

The analyses were performed using cross-sectional data; no causal inferences can be drawn. Besides, self-reporting for several variables may introduce some recall or reporting bias. Third, the BDHS dataset lacked some very key health system indicators, such as facility accessibility, quality of care, and distance

to health services, that could have enriched the analyses even further. Further studies using longitudinal data and more health-system variables should be conducted to identify the determinants influencing skilled birth attendance in Bangladesh.

Abbreviations

SBA (Skilled Birth Assistance)

ANC (Antenatal Care)

PNC (Postnatal care)

ML (Machine Learning)

LMICs (low- and middle-class countries)

WHO (World health Organization)

SDG (Sustainable Development Goal)

BDHS (Bangladesh Demographic and Health Survey)

DHS (Demographic and health Survey)

EAs (Enumeration Areas)

USAID (United States Agency for International Development)

MICE (Multiple Imputation by Chained Equations (MICE)

SMOTE (Synthetic Minority Over-sampling Technique)

SHAP (SHapley Additive exPlanations)

LIME (Local Interpretable Model-diagnostic Explanations)

DT (Decision Tree)

RF (Random Forest)

SVM (Support Vector Machine)

LR (Logistic Regression)

KNN (k-Nearest Neighbors)

XGBoost (Extreme Gradient Boosting)

ANN (Artificial Nural Network)

RBF (Radial Basis Function (kernel for SVM))

AUCROC (Area Under the Receiver Operating Characteristics Curve)

MCC (Matthews Correlation Coefficient)

Kappa (Cohen's Kappa)

STROBE (Strengthening the Reporting of Observational Studies in Epidemiology)

Declarations

Ethical Considerations

Written informed consent was obtained from all participants prior to interviews. Permission to use the anonymized dataset for secondary analysis was granted by the **DHS Program**, and details available at The DHS Program - Protecting the Privacy of DHS Survey Respondents.

Data Availability Statement

The data that support the findings of this study are publicly available from the Bangladesh Demographic and Health Survey (BDHS) 2022. These datasets can be accessed upon reasonable request through the DHS Program website: https://dhsprogram.com.

Declaration of AI Assistance

However, artificial intelligence-driven tools ChatGPT, Perplexity, and QuillBot were utilized only to enhance grammar, wording, and formatting during manuscript preparation. The authors carried out all the data analyses, interpretations as well as conclusions.

CRediT Authorship Contribution Statement

Md Salek Miah (Corresponding Author: Conceptualization, Investigation, Methodology, Data Curation, Formal Analysis, Visualization, Resources, Project Administration, Writing – Original Draft, and Final Manuscript Supervision. **Mst. Labonna Begum & Fhameda Faija Lamia:** Writing **-** Original Draft, Resources, Project Administration.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Consent for publication

Not applicable.

Code Availability

The code supporting the findings of this study is available from the corresponding author upon reasonable request.

Acknowledgement

The authors thank Md. Yousuf Ali (University of Oulu, EMJM Promise – Sustainable Mineral and Metal Processing Engineering) for valuable guidance and suggestions throughout the study.

References

1. Salmi, M., Atif, D., Oliva, D., Abraham, A. & Ventura, S. Handling imbalanced medical datasets: review of a decade of research. *Artif Intell Rev* 57, (2024).
2. Christoffersen, L., Teigen, J. & Rønningstad, C. Following-up midwives after adverse incidents: How front-line management practices help second victims. *Midwifery* 85, (2020).
3. WHO. Maternal health.
4. Koski, P., Raussi-Lehto, E., Leskinen, P. & Klemetti, R. Patient perception of labor support behaviors provided by Finnish midwives. *Midwifery* 131, (2024).
5. Gordon, F., Whitburn, L. Y. & Newton, M. Understanding the impact of Australian Health Practitioner Regulation Agency notifications on midwives' wellbeing, midwifery practise, and career: A qualitative study. *Midwifery* 148, (2025).
6. Memmott, C. *et al.* 'Forgotten as first line providers': The experiences of midwives during the COVID-19 pandemic in British Columbia, Canada. *Midwifery* 113, 103437 (2022).
7. Nahar, M. T., Ibn Anik, S. M. F., Islam, M. A. & Islam, S. M. S. Individual and community-level factors associated with skilled birth attendants during delivery in Bangladesh: A multilevel analysis of demographic and health surveys. *PLoS One* 17, (2022).
8. The Global Goal. Good Health and well being.
9. Chowdhury, S. S. A., Kundu, S. & Sharif, A. Bin. Socioeconomic and geographical inequalities in using skilled birth attendants during delivery in Bangladesh over two decades. *BMC Pregnancy Childbirth* 23, (2023).
10. Mmr, Y. & Pm, †. *Maternal Mortality Bangladesh 2000-2020 Internationally Comparable MMR Estimates by the Maternal Mortality Inter-Agency Group (MMEIG): WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division*.
11. Islam, N., Islam, M. T. & Yoshimura, Y. Practices and determinants of delivery by skilled birth attendants in Bangladesh. *Reprod Health* 11, (2014).

12. Nahar, M. T., Ibn Anik, S. M. F., Islam, M. A. & Islam, S. M. S. Individual and community-level factors associated with skilled birth attendants during delivery in Bangladesh: A multilevel analysis of demographic and health surveys. *PLoS One* 17, (2022).
13. Mahfuzur, M. R. *et al.* Exploring spatial variations in level and predictors of unskilled birth attendant delivery in Bangladesh using spatial analysis techniques: Findings from nationally representative survey data. *PLoS One* 17, (2022).
14. Islam, N., Islam, M. T. & Yoshimura, Y. Practices and determinants of delivery by skilled birth attendants in Bangladesh. *Reprod Health* 11, (2014).
15. Thapa, N. R. Factors Associated with Skilled Birth Attendance in Nepal. *Curriculum Development Journal* 32, 125–141 (2024).
16. *Bangladesh Demographic and Health Survey 2022 Final Report*. www.niport.gov.bd (2024).
17. Ryan, B. L., Krishnan, R. J., Terry, A. & Thind, A. Do four or more antenatal care visits increase skilled birth attendant use and institutional delivery in Bangladesh? A propensity-score matched analysis. *BMC Public Health* 19, (2019).
18. Ryan, B. L., Krishnan, R. J., Terry, A. & Thind, A. Do four or more antenatal care visits increase skilled birth attendant use and institutional delivery in Bangladesh? A propensity-score matched analysis. *BMC Public Health* 19, (2019).
19. Van Den Goorbergh, R., Van Smeden, M., Timmerman, D. & Ben Van Calster. The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression. *Journal of the American Medical Informatics Association* 29, 1525–1534 (2022).
20. Musie, M. R. & Mulaudzi, F. M. Knowledge and attitudes of midwives towards collaboration with traditional birth attendants for maternal and neonatal healthcare services in rural communities in South Africa. *Midwifery* 130, (2024).
21. *Trends in Maternal Mortality Estimates 2000 to 2023*.
22. Mugo, N. S., Agho, K. E. & Dibley, M. J. Risk Factors for Non-use of Skilled Birth Attendants: Analysis of South Sudan Household Survey, 2010. *Matern Child Health J* 20, 1266–1279 (2016).
23. Yakoob, M. Y. *et al.* The effect of providing skilled birth attendance and emergency obstetric care in preventing stillbirths. *BMC Public Health* vol. 11 Preprint at https://doi.org/10.1186/1471-2458-11-S3-S7 (2011).
24. Berhan, Y. & Berhan, A. SKILLED HEALTH PERSONNEL ATTENDED DELIVERY AS A PROXY INDICATOR FOR MATERNAL AND PERINATAL MORTALITY: A SYSTEMATIC REVIEW. https://doi.org/10.4314/ejhs.v24i1.7S doi:10.4314/ejhs.v24i1.7S.
25. Blencowe, H. *et al.* National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: A systematic analysis. *Lancet Glob Health* 4, e98–e108 (2016).
26. Yaya, S. *et al.* Predictors of skilled birth attendance among married women in Cameroon: further analysis of 2018 Cameroon Demographic and Health Survey. *Reprod Health* 18, (2021).

27. Shimeka Teferra, A., Mazengia Alemu, F. & Meseret Woldeyohannes, S. *Institutional Delivery Service Utilization and Associated Factors among Mothers Who Gave Birth in the Last 12 Months in Sekela District, North West of Ethiopia: A Community-Based Cross Sectional Study*. http://www.biomedcentral.com/1471-2393/12/74 (2012).
28. Dong, Y. & Peng, C.-Y. J. *Principled Missing Data Methods for Researchers*. http://www.springerplus.com/content/2/1/222 (2013).
29. Ling, C. X. & Sheng, V. S. *Cost-Sensitive Learning and the Class Imbalance Problem Motivation and Background*. (2008).
30. Sarakshetrin, A., Sujimongkol, C., Rongmuang, D., Chantra, R. & Nimwatanakul, S. Support Vector Machine-Based Prediction Model for Healthcare Workforce Transition Success Under Decentralization. *F1000Res* 14, 49 (2025).
31. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. https://doi.org/10.1145/2939672.2939785 (2016) doi:10.1145/2939672.2939785.
32. Hakkal, S. & Lahcen, A. A. XGBoost To Enhance Learner Performance Prediction. *Computers and Education: Artificial Intelligence* 7, (2024).
33. Zhang, P., Jia, Y. & Shang, Y. Research and application of XGBoost in imbalanced data. *Int J Distrib Sens Netw* 18, (2022).
34. Walle, A. D., Kebede, S. D., Adem, J. B. & Mamo, D. N. Machine-learning algorithm to predict home delivery after antenatal care visit among reproductive age women in East Africa. *Front Glob Womens Health* 6, (2025).
35. Vu, T. *et al.* Prediction of depressive disorder using machine learning approaches: findings from the NHANES. *BMC Med Inform Decis Mak* 25, (2025).
36. Moon, Y. & Woo, H. Key risk factors of generalized anxiety disorder in adolescents: machine learning study. *Front Public Health* 12, (2024).
37. Moon, Y. & Woo, H. Key risk factors of generalized anxiety disorder in adolescents: machine learning study. *Front Public Health* 12, (2024).
38. *Bangladesh Demographic and Health Survey 2022 Final Report*. www.niport.gov.bd (2024).
39. Chang, W., Chen, X., He, Z. & Zhou, S. A Prediction Hybrid Framework for Air Quality Integrated with W-BiLSTM(PSO)-GRU and XGBoost Methods. *Sustainability (Switzerland)* 15, (2023).
40. Li, Y., Yang, Y., Song, P., Duan, L. & Ren, R. An improved SMOTE algorithm for enhanced imbalanced data classification by expanding sample generation space. *Sci Rep* 15, (2025).
41. Fernández, A., García, S., Herrera, F. & Chawla, N. V. *SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary*. *Journal of Artificial Intelligence Research* vol. 61 (2018).
42. Li, Y., Yang, Y., Song, P., Duan, L. & Ren, R. An improved SMOTE algorithm for enhanced imbalanced data classification by expanding sample generation space. *Sci Rep* 15, (2025).

43. Johnsen, H., Christensen, U., Juhl, M. & Villadsen, S. F. Implementing the MAMAACT intervention in Danish antenatal care: a qualitative study of non-Western immigrant women's and midwives' attitudes and experiences. *Midwifery* 95, (2021).
44. Khumoetsile Daphney, S., Tinda, R., Antoinette, D. P. & Petra, B. Experiences of midwives regarding provision of culturally competent care to women receiving maternal care in South Africa. *Midwifery* 116, (2023).
45. Leavy, F. & Leggett, H. Midwives' experiences of working in team continuity of carer models: A qualitative evidence synthesis. *Midwifery* 112, (2022).
46. Islam, A. *et al.* Prevalence and determinants of utilizing skilled birth attendance during home delivery of pregnant women in India: Evidence from the Indian Demographic and Health Survey 2015–16. *PLoS One* 19, (2024).
47. Budu, E. *et al.* Early age at first childbirth and skilled birth attendance during delivery among young women in sub-Saharan Africa. *BMC Pregnancy Childbirth* 21, (2021).
48. Islam, A. *et al.* Prevalence and determinants of utilizing skilled birth attendance during home delivery of pregnant women in India: Evidence from the Indian Demographic and Health Survey 2015–16. *PLoS One* 19, (2024).
49. Walle, A. D., Kebede, S. D., Adem, J. B. & Mamo, D. N. Machine-learning algorithm to predict home delivery after antenatal care visit among reproductive age women in East Africa. *Front Glob Womens Health* 6, (2025).
50. Zhao, Y., Wong, Z. S. Y. & Tsui, K. L. A Framework of Rebalancing Imbalanced Healthcare Data for Rare Events' Classification: A Case of Look-Alike Sound-Alike Mix-Up Incident Detection. *J Healthc Eng* 2018, (2018).
51. Pérez-Castejón, M., Suárez-Cortés, M., Jiménez-Ruiz, I. & Jiménez-Barbero, J. A. Job satisfaction in midwives working in labour ward: A systematic review with meta-analysis. *Midwifery* vol. 137 Preprint at https://doi.org/10.1016/j.midw.2024.104112 (2024).
52. Olawade, D. B. *et al.* Determinants of maternal mortality in south-western Nigeria: Midwives' perceptions. *Midwifery* 127, (2023).
53. Chowdhury, S. S. A., Kundu, S. & Sharif, A. Bin. Socioeconomic and geographical inequalities in using skilled birth attendants during delivery in Bangladesh over two decades. *BMC Pregnancy Childbirth* 23, (2023).
54. Alkema, L. *et al.* Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the un Maternal Mortality Estimation Inter-Agency Group. *The Lancet* 387, 462–474 (2016).

Tables

Table 1:Summary of Machine Learning Models, Hyperparameters, and Their Description Used for Predicting Skilled Birth Attendance

Model	Parameter	Value	Description
Random Forest	mtry	2, 4, 6, 8	Number of variables randomly sampled at each split in a tree
Decision Tree	cp	0.001, 0.011, 0.021, 0.031, 0.041, 0.051	Complexity parameter controlling tree pruning
K-Nearest Neighbors (KNN)	k	3, 5, 7, 9, 11, 13, 15	Number of nearest neighbors used for classification
Logistic Regression	-	No tuning needed	Linear model for binary outcome; no hyperparameters
Support Vector Machine (SVM)	sigma	0.01	RBF kernel parameter controlling influence of individual points
	C	0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2	Penalty parameter controlling misclassification cost
XGBoost	nrounds	100, 200	Number of boosting iterations (trees)
	max_depth	3, 5, 7	Maximum depth of each tree
	eta	0.01, 0.1, 0.3	Learning rate (step size for weight updates)
	gamma	0	Minimum loss reduction required to make a split
	colsample_bytree	0.8	Fraction of features sampled for each tree
	min_child_weight	1	Minimum sum of instance weights required in a child node
	subsample	0.8	Fraction of rows sampled for building each tree
Neural Network	size	5, 10, 15	Number of neurons in the hidden layer
	decay	0, 0.01, 0.1	Weight decay regularization to prevent overfitting

Table 2: Characteristics of Women by Skilled Birth Attendance (SBA), Bangladesh

Target Variables	Skilled	Unskilled	Total n (%)	P-value
	n (%)	n (%)		
SBA	1,479 (30.36)	3,392(69.64)	4, 871 (100)	-
Predictors				
Demographic Variables				
Maternal Age				0.002
<20	460 (13.6)	200 (13.5)	660 (13.6)	
20-34	2,672 (78.8)	1,120 (75.7)	3,792 (77.9)	
>35	260 (7.7)	159 (10.8)	419 (8.6)	
Maternal age at first birth				**<0.001**
<20 years	1,877 (55.3)	1,034 (69.9)	2,911 (59.8)	
≥20 years	1,515 (44.7)	445 (30.1)	1,960 (40.2)	
Age at first sex				**<0.001**
<18 years	1,912 (56.4)	1,013 (68.5)	2,925 (60.1)	
≥18 years	1,480 (43.6)	466 (31.5)	1,946 (39.9)	
Maternal education				**<0.001**
None	114 (3.4)	131 (8.9)	245 (5.0)	
Primary	593 (17.5)	518 (35.0)	1,111 (22.8)	
Secondary+	2,685 (79.2)	830 (56.1)	3,515 (72.2)	
Husband Education				**<0.001**
No education	354 (10.4)	336 (22.7)	690 (14.2)	
Basic	835 (24.6)	586 (39.6)	1,421 (29.2)	
Secondary	2,167 (63.9)	539 (36.4)	2,706 (55.6)	
Don’t know	36 (1.1)	18 (1.2)	54 (1.1)	
Household religion				**<0.001**
Muslim	3,044 (89.7)	1,415 (95.7)	4,459 (91.5)	
Non-Muslim	348 (10.3)	64 (4.3)	412 (8.5)	
Household & Socioeconomic Variables				
Residence				**<0.001**

Urban	1,285 (37.9)	332 (22.5)	1,617 (33.2)	
Rural	2,107 (62.1)	1,147 (77.6)	3,254 (66.8)	
Wealth category				**<0.001**
Poor	1,055 (31.1)	901 (60.9)	1,956 (40.2)	
Middle	689 (20.3)	281 (19.0)	970 (19.9)	
Rich	1,648 (48.6)	297 (20.1)	1,945 (39.9)	
Children category				**<0.001**
Single	1,470 (44.0)	413 (28.4)	1,883 (39.3)	
Multiple	1,875 (56.0)	1,040 (71.6)	2,915 (60.8)	
Household size				0.054
≤5 members	1,881 (55.4)	776 (52.5)	2,657 (54.6)	
>5 members	1,511 (44.6)	703 (47.5)	2,214 (45.4)	
Health insurance				0.028
Yes	16 (0.5)	1 (0.1)	17 (0.4)	
No	3,376 (99.5)	1,478 (99.9)	4,854 (99.7)	
Decision-making				0.009
Yes	1,890 (56.2)	763 (52.2)	2,653 (55.0)	
No	1,472 (43.8)	700 (47.8)	2,172 (45.0)	
Household assets				**<0.001**
Yes	2,177 (64.2)	689 (46.6)	2,866 (58.8)	
No	1,215 (35.8)	790 (53.4)	2,005 (41.2)	
Household materials				0.001
Improved	2,939 (86.7)	1,330 (89.9)	4,269 (87.6)	
Unimproved	453 (13.4)	149 (10.1)	602 (12.4)	
Obstetric & Reproductive Variables				
Parity				**<0.001**
Primiparo	1,437 (42.6)	399 (27.3)	1,836 (38.0)	
Multiparo	1,937 (57.4)	1,064 (72.7)	3,001 (62.0)	
ANC visits (≥4)				**<0.001**

Yes	1,684 (49.6)	328 (22.2)	2,012 (41.3)	
No	1,708 (50.4)	1,151 (77.8)	2,859 (58.7)	
Delivery mode				**<0.001**
Vaginal	1,134 (33.5)	1,462 (98.9)	2,596 (53.4)	
Caesarean	2,249 (66.5)	16 (1.1)	2,265 (46.6)	
Birth size				0.409
Small/Under	483 (14.2)	224 (15.2)	707 (14.5)	
Average/Other	2,909 (85.8)	1,255 (84.8)	4,164 (85.5)	
PNC				**<0.001**
Yes	3,080 (90.8)	1,031 (69.7)	4,111 (84.4)	
No	312 (9.2)	448 (30.3)	760 (15.6)	
Pregnancy pressure				0.859
Pressured	109 (3.2)	46 (3.1)	155 (3.2)	
Not pressured	3,253 (96.8)	1,417 (96.9)	4,670 (96.8)	
Sexually active				0.004
Yes	2,614 (77.1)	1,194 (80.7)	3,808 (78.2)	
No	778 (22.9)	285 (19.3)	1,063 (21.8)	
Fertility preference				**<0.001**
Want more	1,853 (54.6)	702 (47.5)	2,555 (52.4)	
Do not want	1,539 (45.4)	777 (52.5)	2,316 (47.6)	
Pregnancy loss				0.005
Yes	707 (20.8)	257 (17.4)	964 (19.8)	
No	2,685 (79.2)	1,222 (82.6)	3,907 (80.2)	
Preterm birth				**<0.001**
Preterm	327 (9.6)	89 (6.0)	416 (8.5)	
Term	3,065 (90.4)	1,390 (94.0)	4,455 (91.5)	
Multiple pregnancy				0.079
Yes	45 (1.3)	11 (0.7)	56 (1.1)	
No	3,347 (98.7)	1,468 (99.3)	4,815 (98.9)	

Currently pregnant				**<0.001**
Yes	101 (3.0)	75 (5.1)	176 (3.6)	
No	3,291 (97.0)	1,404 (94.9)	4,695 (96.4)	
Media Exposure & Empowerment				
Media exposure				**<0.001**
Any	3,024 (89.2)	1,104 (74.7)	4,128 (84.8)	
None	368 (10.8)	375 (25.3)	743 (15.3)	
Internet use				**<0.001**
Yes	1,327 (39.1)	270 (18.3)	1,597 (32.8)	
No	2,065 (60.9)	1,209 (81.7)	3,274 (67.2)	
Geographic & Contextual Variables				
Division				**<0.001**
Barishal	357 (10.5)	181 (12.2)	538 (11.0)	
Chattogram	571 (16.8)	284 (19.2)	855 (17.6)	
Dhaka	535 (15.8)	182 (12.3)	717 (14.7)	
Khulna	489 (14.4)	77 (5.2)	566 (11.6)	
Mymensingh	344 (10.1)	250 (16.9)	594 (12.2)	
Rajshahi	380 (11.2)	116 (7.8)	496 (10.2)	
Rangpur	373 (11.0)	176 (11.9)	549 (11.3)	
Sylhet	343 (10.1)	213 (14.4)	556 (11.4)	
Toilet type				**<0.001**
Improved	2,548 (75.1)	952 (64.4)	3,500 (71.9)	
Unimproved	844 (24.9)	527 (35.6)	1,371 (28.2)	
Drinking water				0.001
Improved	2,884 (85.0)	1,312 (88.7)	4,196 (86.1)	
Unimproved	508 (15.0)	167 (11.3)	675 (13.9)	
IPV				0.002
Yes	368 (11.1)	209 (14.3)	577 (12.1)	
No	2,951 (88.9)	1,249 (85.7)	4,200 (87.9)	

Table 3:Performance Metrics of Machine Learning Models

Model	Accuracy	Precision	Recall	F1	MCC	Kappa	AUROC
Random Forest	0.82	0.90	0.84	0.87	0.60	0.59	0.91
Decision Tree	0.81	0.91	0.80	0.85	0.59	0.58	0.88
KNN	0.79	0.91	0.77	0.83	0.55	0.54	0.86
Logistic Regression	0.82	0.96	0.77	0.86	0.65	0.62	0.90
SVM	0.80	0.98	0.74	0.84	0.64	0.60	0.90
XGBoost	0.84	0.91	0.85	0.88	0.63	0.63	0.91
Neural Network	0.80	0.91	0.80	0.85	0.58	0.57	0.89

Figures

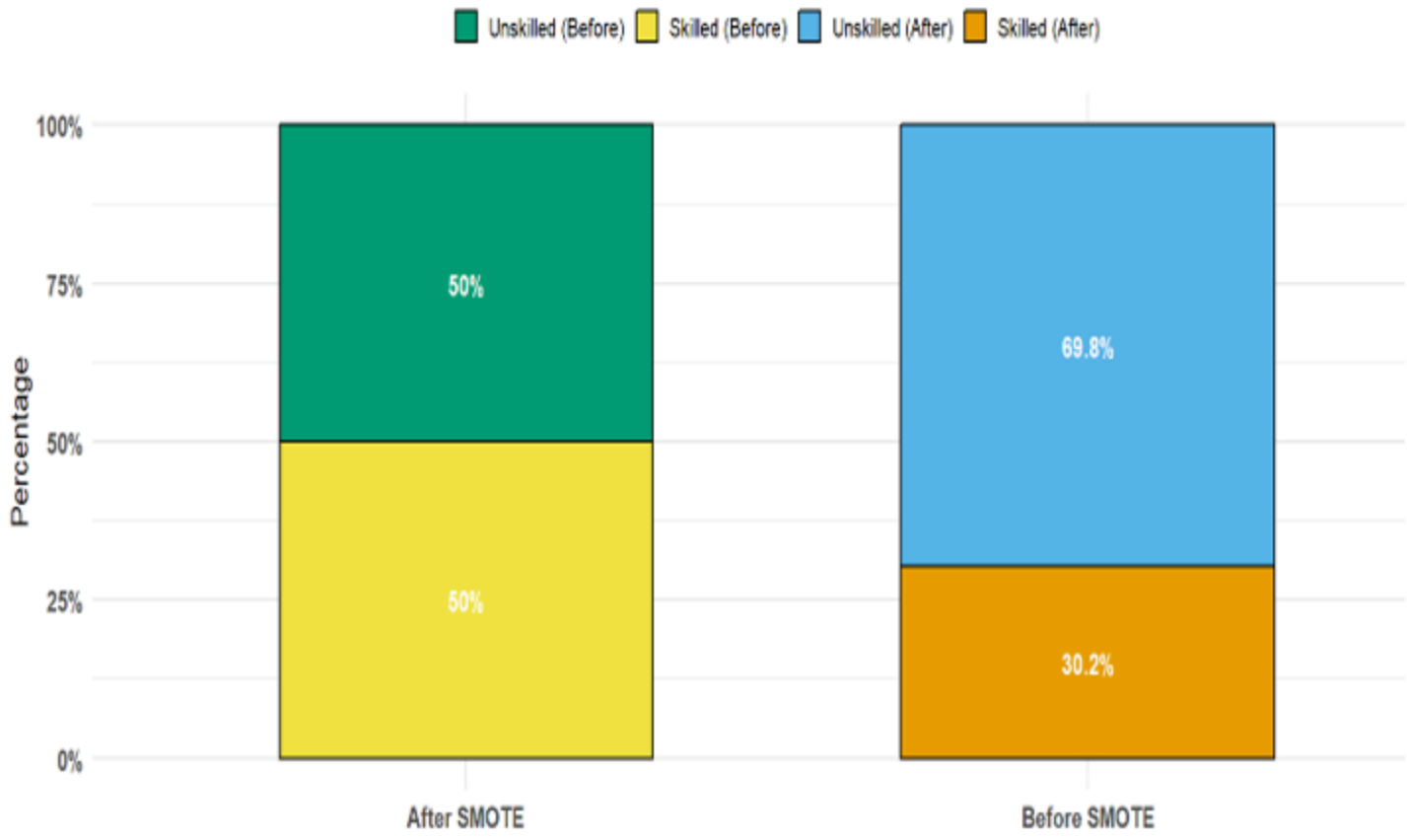


Figure 1

Effect of SMOTE on Outcome Class Balance

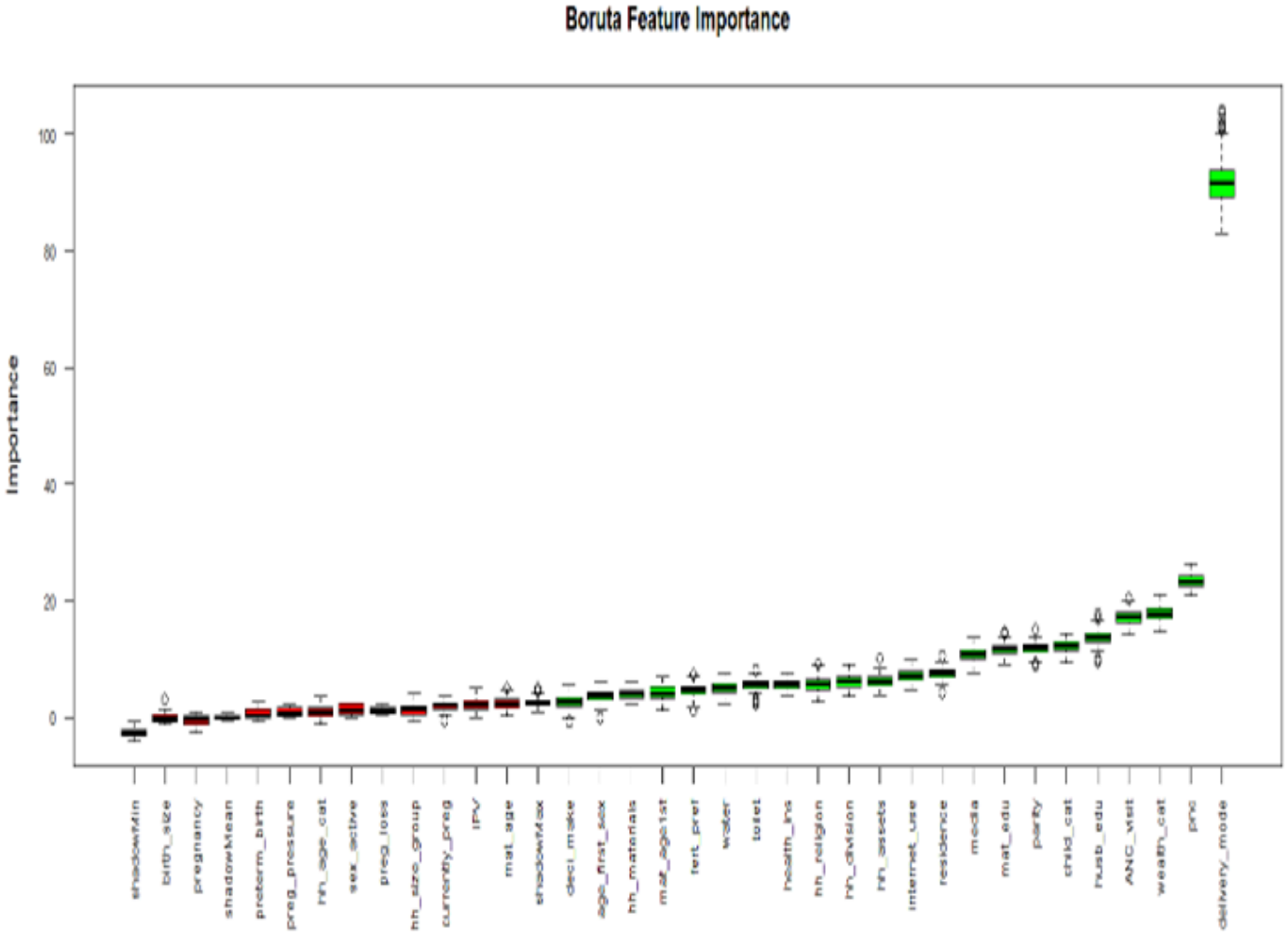


Figure 2

Feature Selection using Boruta Algorithm

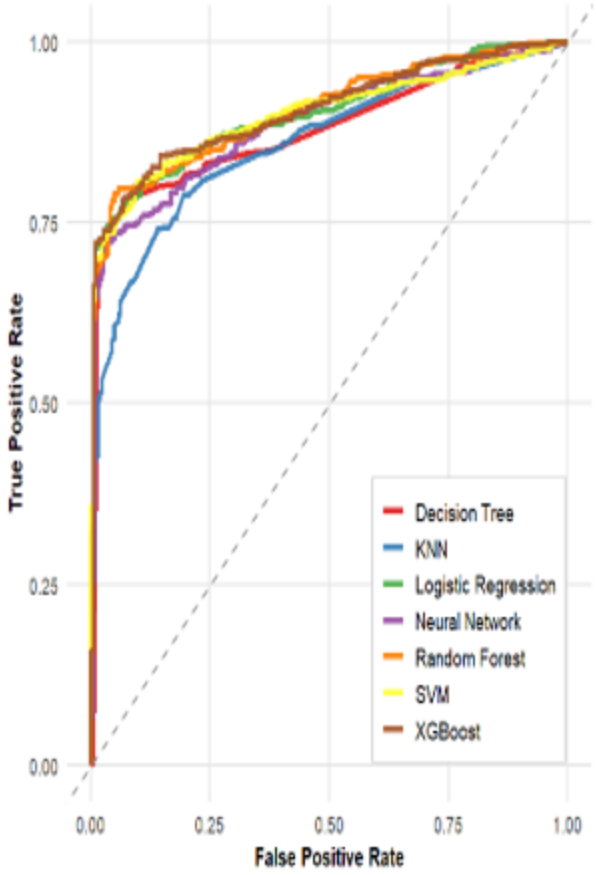


Figure 3

Comparative ROC Curves of ML Classifiers

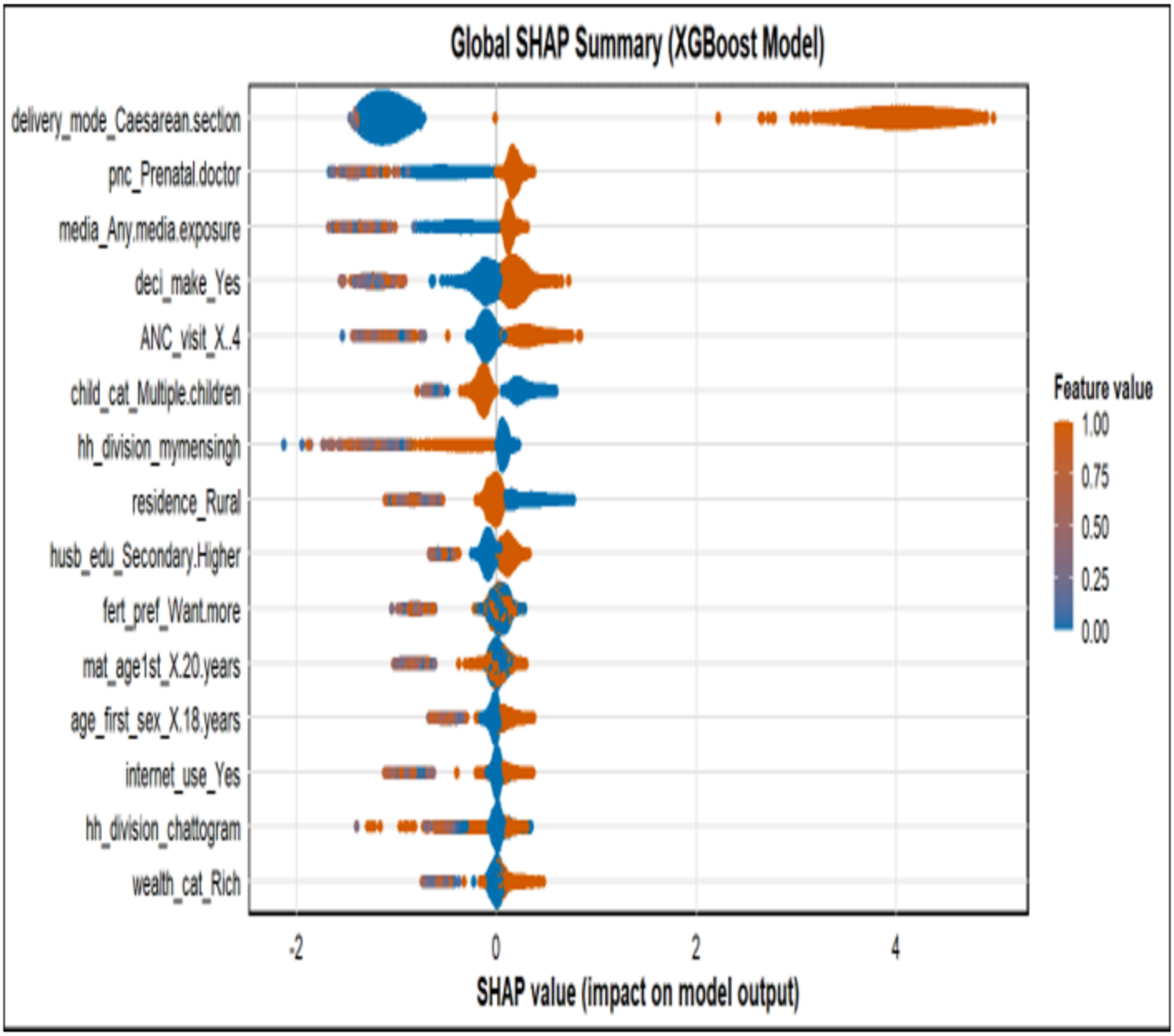


Figure 4

Feature Importance for Skilled Birth Assistance Predicted by SHAP Values

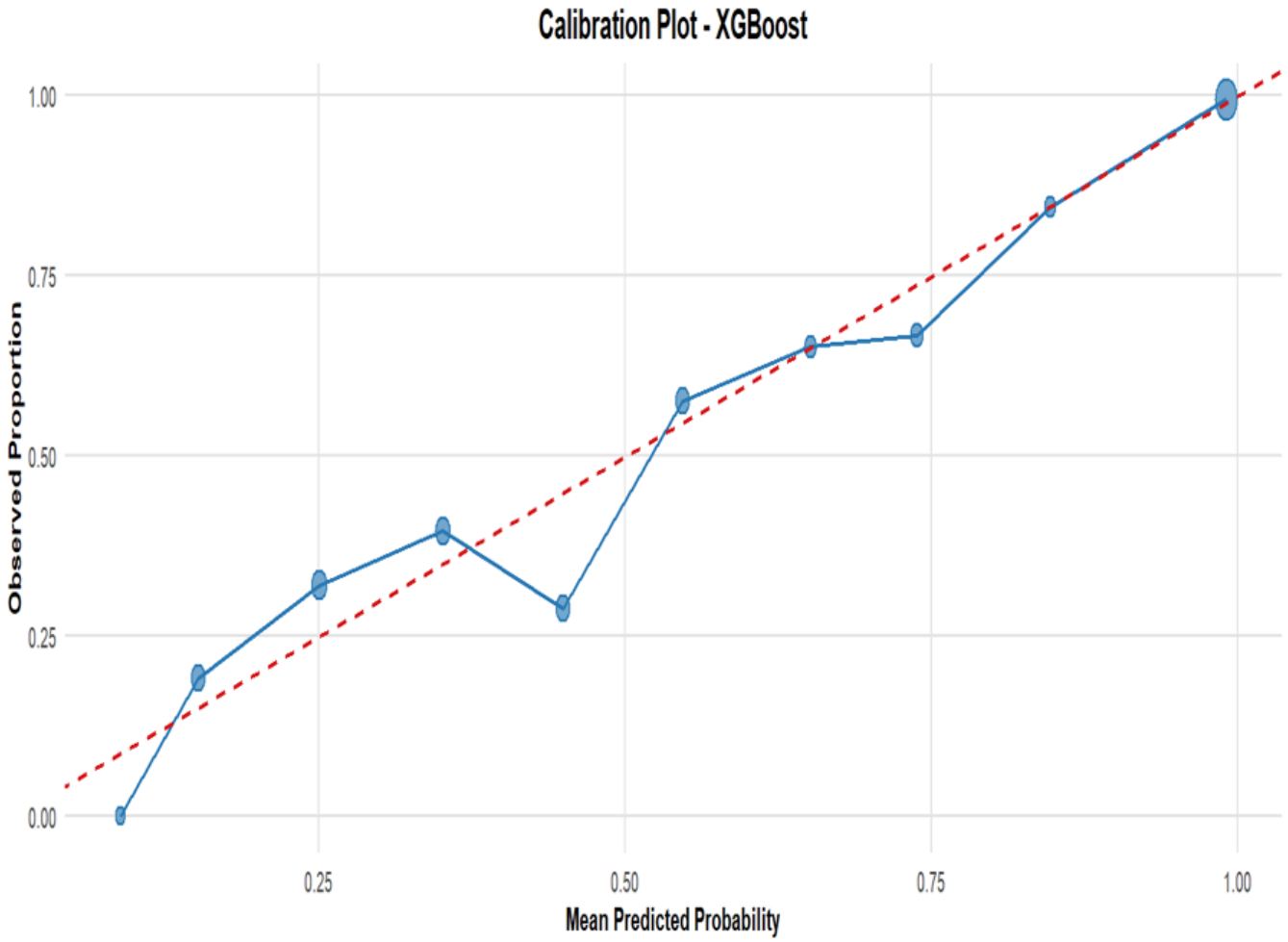


Figure 5

XGBoost Model Calibration Curve

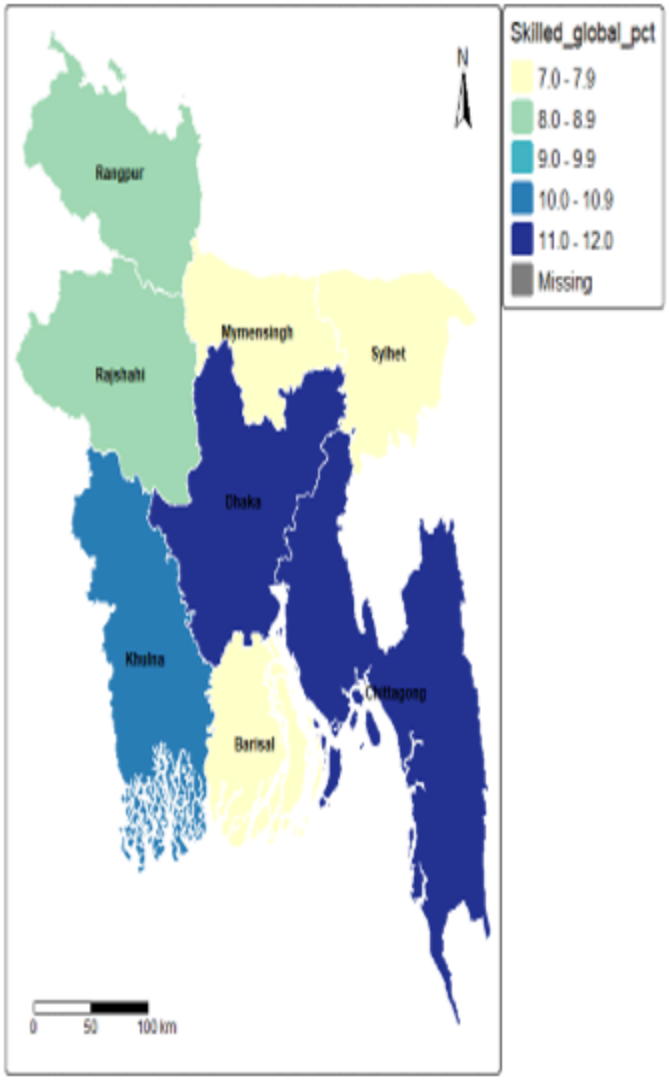


Figure 6

Spatial Variation of SBA

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureS1.png
- SupplymentaryFigureS2.tiff
- SupplymentaryfigureS3.tiff
- SupplymentaryFigureS4.tiff