

RESEARCH

Open Access



Predicting adverse birth outcome among childbearing women in Sub-Saharan Africa: employing innovative machine learning techniques

Habtamu Setegn Ngusie^{1*}, Shegaw Anagaw Mengiste², Alemu Birara Zemariam³, Bogale Molla⁴, Getanew Aschalew Tesfa⁵, Binyam Tariku Seboka⁵, Tilahun Dessie Alene⁶ and Jing Sun^{7,8}

Abstract

Background Adverse birth outcomes, including preterm birth, low birth weight, and stillbirth, remain a major global health challenge, particularly in developing regions. Understanding the possible risk factors is crucial for designing effective interventions for birth outcomes. Accordingly, this study aimed to develop a predictive model for adverse birth outcomes among childbearing women in Sub-Saharan Africa using advanced machine learning techniques. Additionally, this study aimed to employ a novel data science interpretability techniques to identify the key risk factors and quantify the impact of each feature on the model prediction.

Methods The study population involved women of childbearing age from 26 Sub-Saharan African countries who had given birth within five years before the data collection, totaling 139,659 participants. Our data source was a recent Demographic Health Survey (DHS). We utilized various data balancing techniques. Ten advanced machine learning algorithms were employed, with the dataset split into 80% training and 20% testing sets. Model evaluation was conducted using various performance metrics, along with hyperparameter optimization. Association rule mining and SHAP analysis were employed to enhance model interpretability.

Results Based on our findings, about 28.59% (95% CI: 28.36, 28.83) of childbearing women in Sub-Saharan Africa experienced adverse birth outcomes. After repeated experimentation and evaluation, the random forest model emerged as the top-performing machine learning algorithm, with an AUC of 0.95 and an accuracy of 88.0%. The key risk factors identified were home deliveries, lack of prenatal iron supplementation, fewer than four antenatal care (ANC) visits, short and long delivery intervals, unwanted pregnancy, primiparous mothers, and geographic location in the West African region.

Conclusion The region continues to face persistent adverse birth outcomes, emphasizing the urgent need for increased attention and action. Encouragingly, advanced machine learning methods, particularly the random forest algorithm, have uncovered crucial insights that can guide targeted actions. Specifically, the analysis identifies

*Correspondence:
Habtamu Setegn Ngusie
habtamuhi3@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

risky groups, including first-time mothers, women with short or long birth intervals, and those with unwanted pregnancies. To address the needs of these high-risk women, the researchers recommend immediately providing iron supplements, scheduling comprehensive prenatal care, and strongly encouraging facility-based deliveries or skilled birth attendance.

Keywords Advanced machine learning algorithm, Association rule mining, Adverse birth outcome, Sub-saharan Africa

Introduction

In recent years, there has been growing recognition of the profound impact that adverse birth outcomes can have on the long-term health and well-being of individuals, spanning from infancy to adulthood [1, 2]. Adverse birth outcomes, including preterm birth, stillbirth, macrosomia, and low birth weight, occur during the period of fetal development and the early weeks of life [3].

While considerable progress has been made in reducing adverse birth outcomes globally, substantial disparities persist between developing and developed countries, with Sub-Saharan Africa bearing a hefty burden [4–7]. Approximately 30 million low-birth-weight babies are born worldwide each year, which accounts for around 23.4% of all births [8, 9]. Low birth weight poses various adverse health risks for the baby, like fetal and neonatal mortality, morbidity, and impaired growth and cognitive development, as well as leading to chronic diseases later in life [10].

Sub-Saharan Africa and Southeast Asia were the hotspots of low birth weight, which shared three-quarters of the global burden in 2020. Sub-Saharan Africa alone shares 27.1% of the worldwide burden of low birth weight [9]. Both low birth weight and macrosomia or high birth weight are critical indicators of infant health. According to the DHS report, the pooled prevalence of macrosomia in this region was 8.33% [11].

Prematurity is also a major cause of neonatal mortality worldwide, with approximately 15 million babies born preterm each year, accounting for around 11% of all deliveries globally [12]. It accounts for 35% of all deaths among newborns worldwide [7]. According to recent estimates, about 65% of preterm births were from Southern Asia and Sub-Saharan Africa [6]. In 2019, it was estimated that around 2.0 million babies were stillborn worldwide. The global stillbirth rate was approximately 13.9 stillbirths per 1,000 total births. Sub-Saharan Africa was the top vulnerable region, which shares 43.6% of the worldwide burden of stillbirths [13].

The health sector's main focus in the Sustainable Development Goals (SDGs) is reducing infant and child mortality, which is closely linked to adverse birth outcomes [14]. Alleviating adverse birth outcomes is vital for tackling infant and childhood mortality and in turn, achieving the SDGs. Various strategies, including reducing adverse birth outcomes like low birth weight, stillbirth, and

preterm births a part of the SDGs, have been undertaken by different stakeholders to address the magnitude of these issues [15]. Despite these initiatives, adverse birth outcomes remain a significant challenge in Sub-Saharan Africa, with an overall pooled prevalence of 29.7% based on previous DHS data [16].

Previous literature highlighted several possible factors for adverse birth outcomes, including the place of residence [17–21], maternal education [3, 16, 19, 20, 22, 23], mode of delivery [22], substance use [22], distance to healthcare facilities [16], gender of the child [3, 16, 17, 19, 24, 25], maternal age [3, 18, 26–29], religion [20], number of children [24], wealth index [16, 19–22, 25], parity [16–19, 22, 27, 29], contraceptive use history [30], marital status [3, 21, 22], husband's education level [3, 24], being anemic [31], iron supplementation [20, 24, 31], type of pregnancy [3, 16, 21, 24], number of ANC visits [16, 18, 23, 25, 26, 30, 32], birth interval [3, 21, 30, 33], unwanted pregnancy [22], place of delivery [3, 20, 29], and geographic region [19, 22, 34].

While previous studies have examined the factors influencing adverse birth outcomes using DHS data from various countries in Sub-Saharan Africa [16, 28, 33, 35], a deeper understanding requires the use of advanced machine learning algorithms and data science techniques. This approach enables the discovery of hidden patterns and relationships that may not be easily identified through traditional statistical methods. Additionally, our study's use of recent DHS data allows us to capture the most up-to-date trends and factors influencing adverse birth outcomes in the region.

Accordingly, we proposed a study aimed at developing a predictive model for adverse birth outcomes among childbearing women in Sub-Saharan Africa using advanced machine learning techniques. Additionally, this study aimed to employ novel data science interpretability techniques to identify the key risk factors and quantify the impact of each feature on the model prediction.

Method

Data source

Secondary data from the most recent DHS, which was carried out in 26 countries in Sub-Saharan Africa, were used in this analysis. The following are the nations and corresponding survey years: Angola (2015–2016), Benin (2017–2018), Burkina Faso (2021), Burundi (2016–17),

Cameroon (2018), Ivory Coast (2021), Ethiopia (2016), Gabon (2019–21), Gambia (2019–20), Ghana (2022), Guinea (2018), Kenya (2022), Liberia (2019–20), Madagascar (2021), Malawi (2015–16), Mali (2018), Mauritania (2019–21), Nigeria (2018), Rwanda (2019–20), Senegal (2019), Sierra Leone (2019), South Africa (2016), Tanzania (2022), Uganda (2016), Zambia (2018), and Zimbabwe (2015). The data for this study was obtained from the official database of the DHS program, which can be accessed at the following URL: <https://dhsprogram.com/data/available-datasets.cfm>.

Standardized cross-sectional surveys were conducted by the DHS Program in more than 90 countries to collect thorough and representative data on nutrition, HIV, health, and population. A multi-stage stratified sampling approach was used in these surveys to choose respondents from households within specific clusters. To select the enumeration areas, sampling strata based on the urban and rural sectors were created, and probability proportionate to size was used in the process. Households were picked using an equal probability systematic sampling procedure within the designated enumeration areas. Detailed information regarding the sampling procedure and methodology employed in the survey can be accessed on the DHS website and their provided guidelines [36].

The study focused on women of childbearing age [15–49] in Sub-Saharan African countries who had given birth within the five years prior to data collection. The analysis encompassed a significant sample size of 139,659 women who met the specified criteria across the 26 countries included in the study. The dataset used in this study contained 28 different features that were considered during the analysis.

Study variables and measurements

Adverse birth outcomes, which were defined as the incidence of one or more particular conditions during recent pregnancies, such as low birth weight, macrosomia, pre-term birth, or stillbirth, were the main measures utilized in this study. This definition was based on the availability of possible items on DHS. Low birth weight, macrosomia, stillbirth, and gestational age of fewer than 37 weeks were all integrated by the researchers to form a single variable. If there were any reported adverse birth outcomes, this variable was coded as 1, indicating that there were adverse birth outcomes, and 0 if none were recorded. The previous study provided detailed information on how this variable was recoded from the DHS dataset [33].

Following a review of the literature, the study looked at several independent variables. The variables that were included in this study were: the place of residence [17–21], maternal education [3, 16, 19, 20, 22, 23], gender of the household head [16, 22], ownership of a mobile phone

[17–21], mode of delivery (cesarean section or vaginal) [22], smoking [22], distance to healthcare facilities [16], maternal employment status [16, 22], gender of the child [3, 16, 17, 19, 24, 25], maternal age [3, 18, 26–29], religion [20], number of children in the family [24], wealth index [16, 19–22, 25], parity (number of previous pregnancies) [16–19, 22, 27, 29], contraceptive use history [30], marital status [3, 20, 29, 30], husband's education level [3, 20, 24, 29], iron supplementation [20, 24, 31], type of pregnancy (single or twin) [3, 16, 21, 24], number of ANC visits [16, 18, 23, 25, 26, 30, 32], birth interval [3, 21, 30, 33], desired pregnancy status (wanted or unwanted pregnancy) [22], first ANC check-up (early or late) [16, 22], place of delivery [3, 20, 29], and media exposure [20, 24, 30, 31].

The birth interval variable was recoded into short (less than 24 months), normal (24 to 59 months), and long (greater than 59 months) categories based on the World Health Organization (WHO)'s guidelines for inter-pregnancy interval [37]. Regarding the timing of first ANC check-up, mothers who started their first ANC visit during the first trimester were considered as early ANC check-ups, and those who started in the second and third trimesters were considered late ANC check-ups [38].

In addition, community-level variables such as country and region (including sub-regions of Sub-Saharan Africa such as West, East, North, Central, and South Africa) were included after repeated experimentation to assess the potential significance of those features in the study. The inclusion of community-level variables, such as country and region, was informed by a review of previous literature [19, 22, 34].

To further assess the importance of these variables, the researchers conducted a series of experiments. Specifically, we included the community-level variables one by one and evaluated the performance of the different machine learning algorithms to determine the impact of these factors on the prediction of adverse birth outcomes. By systematically testing the contribution of the community-level variables, the researchers were able to build a more comprehensive and robust predictive model that accounted for both individual-level and contextual factors.

Data preprocessing

Data pre-processing, the first step in machine learning, is preparing the data to make sure it is suitable for computer analysis [39]. For our models, we used a continuous improvement methodology in our machine learning workflow. We carried out a number of activities, such as feature selection, data balancing, model selection, training, evaluation, parameter optimization via cross-validation, final model selection, and deployment to forecast adverse birth outcomes [40].

Throughout this iterative approach, we continuously improved our models to achieve better outcomes. While Fig. 1 shows the high-level workflow phases visually, it may not capture all the repetitive tasks we completed as part of this process. Beyond the activities depicted in the figure, we engaged in numerous additional iterative steps to systematically refine our models.

The iterative tasks included trying different feature engineering methods to enhance the model inputs, experimenting with various data balancing techniques to address class imbalances, evaluating the performance of different kernel functions for the support vector machine (SVM) model, exploring various model tuning and hyperparameter optimization approaches, and assessing the impact of different n-fold cross-validation schemes.

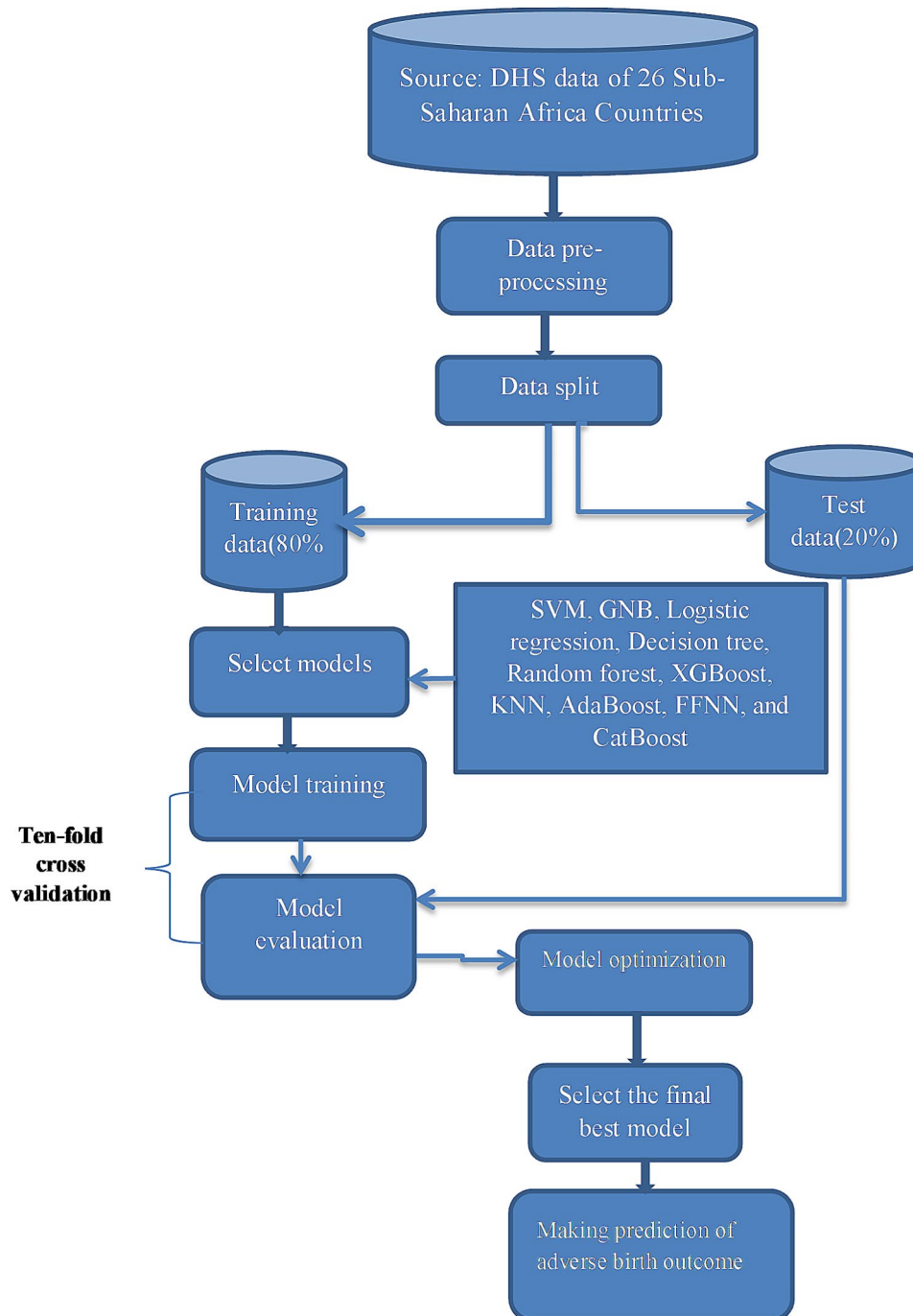


Fig. 1 Study work flow diagram

By continuously iterating and improving our models through this cyclical process, we were able to enhance the accuracy of our adverse birth outcome forecasts.

Moreover, model evaluation is where the trained model is evaluated on the held-out test set data to get an unbiased estimate of its performance, and this was done before model optimization. However, the researcher of this study also conducted model evaluation after optimization and to select the final best model by comparing the optimized models. By continuously iterating and improving our models through this cyclical process, we were able to enhance the accuracy of our adverse birth outcome forecasts.

Data cleaning

Throughout the data analysis process, we ensured the quality and dependability of the dataset by using a rigorous approach. We began with a comprehensive manual audit to identify and eliminate any extraneous data entries. The DHS guideline [36] was used to identify extraneous data entries and outliers. In this study, we identified a few outliers and removed them after assessing the impact through sensitivity analysis [41, 42].

Upon conducting our analysis, we discovered that for each variable utilized in this study, the missing rate was under 10%. To deal with these missing values, we employed the imputation technique known as K-Nearest Neighbors (KNN). A popular method for filling in the missing values is KNN imputation, which uses information from neighboring data points. Because KNN imputation can handle a wide range of data formats, incorporate contextual information, preserve the dataset's structure, and consistently resolve missing values, it is the method of choice for us [43, 44].

Additionally, the correlation matrix was used to assess multicollinearity; a correlation between two variables was considered high if it was greater than 0.8 [45, 46]. Nevertheless, there was no indication of multicollinearity between the variables in our research.

Feature engineering

In order to create precise and useful machine learning models, the process of feature engineering involves carefully locating, extracting, and modifying pertinent features from the given data [47]. We conducted a study using different encoding algorithms for different types of category variables. Specifically, we employed one-hot encoding for nominal categorical variables and label encoding for ordinal categorical variables [48].

The rationale for choosing these encoding methods was based on the nature of the categorical variables in the dataset and the common best practices in the field of machine learning. One-hot encoding and label encoding are well-established techniques that are often the

starting point for handling categorical variables, as they are straightforward to implement and can provide meaningful representations for the machine learning models [49, 50].

Dimensionality reduction

We used multiple dimensionality reduction techniques to improve the performance of the model. Boruta-based feature selection, principal component analysis (PCA), lasso regression, random forest feature selection, recursive feature elimination (RFE), and univariate selection were a few of these techniques. After doing rigorous testing and comparing the outcomes across multiple feature selection techniques, we discovered that RFE and random forest-based feature selection were the most effective strategies in terms of accuracy and robustness.

In our experimentation, we systematically evaluated the performance of various machine learning algorithms before and after applying each feature selection technique. We also compared the performance of the machine learning algorithms across each feature selection technique. This allowed us to determine the optimal dimensionality reduction strategy based on the resulting model improvements.

RFE is an iterative technique that methodically eliminates features from the dataset based on their significance or ability to enhance the model's performance. The less important elements are gradually removed from the model once it has been trained with all of the characteristics up to a predetermined threshold or level of performance [51]. On the other hand, random forest-based feature selection determines which features are most relevant by evaluating each one's relevance using a random forest method.

We were able to remove two characteristics with these methods: smoking and the number of children in the family. Both recursive feature elimination and random forest-based feature selection showed that these traits were the least significant factors.

After removing the two variables, the performance of each machine learning algorithm showed a slight improvement compared to applying other feature selection techniques or no dimensionality reduction at all. For instance, the performance of random forest improved from an accuracy of 83.0% and AUC of 92.0% before feature selection, to an accuracy of 87.0% and AUC of 95.0% after removing the two variables using RFE and random forest-based selection. Additional Material 1 provides a detailed performance comparison of each algorithm before and after feature selection.

Data balancing

In our study, we used a novel data balancing technique to address class imbalance in our dataset. Six data balancing

strategies were used in this approach: the NearMiss algorithm, SMOTE-ENN (SMOTE edited nearest neighbor), Adaptive Synthetic (ADASYN), under and over-sampling, and synthetic minority over-sampling methodology (SMOTE).

We carried out a comprehensive analysis of model performance and took into account several performance parameters to identify the best data balancing method. Using several balancing approaches, we trained machine learning algorithms on the balanced data and compared their performance. We determined the best strategy for resolving the class imbalance in our dataset by this meticulous method [52, 53].

Model selection and development

We select ten advanced machine learning models that could accurately predict adverse birth outcomes. The Jupyter Notebook and the Python packages from sci-kit-learn version 1.3.2 were used in this study. These machine learning algorithms were selected because they could be used for classification tasks and they aligned well with the features of our dataset [54–56].

We used the following algorithms: AdaBoost, eXtreme Gradient Boosting (XGBoost), CatBoost Classifier, Gaussian Naive Bayes (GNB), logistic regression, decision tree, random forest, SVM, KNN method, and feedforward neural network (FFNN). For the detailed descriptions of each employed machine learning algorithm see Additional Material 2.

Model training and evaluation

We used a simple strategy in our study to split the data into two sets: a training set consisting of 80% (111,727 cases) and a testing set consisting of 20% (27,932 cases). The rationale behind choosing a 20% test and 80% train split was based on a careful consideration of the characteristics of our dataset, the complexity of the models employed, and the specific requirements of our project. To further validate the appropriateness of this split, we also compared the performance of each employed algorithm when using a 30% test and 70% train split. The results showed that the 20% test and 80% train split gave a comparable good performance, and we considered this split to be the most suitable for our study [57, 58].

We employed a variety of metrics, including accuracy, precision, recall/sensitivity, F1-score, specificity, and AUC to evaluate each predictive model's performance.

1. Accuracy: The accuracy of a model measures the overall predictions which is determined by dividing the total number of predictions by the number of accurate predictions [59].
2. Precision: Precision assesses how well the model predicts good outcomes. It measures the percentage

of actual positive predictions that match the total number of projected positives. Precision can be calculated using the following formula: $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$, where TP stands for true positive and FP for false positive [59].

3. Recall: Recall evaluates the model's capacity to detect every positive instance. It is also referred to as sensitivity or true positive rate. It calculates the percentage of actual positives that are true positive forecasts [59]. $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$ is the recall formula, in which TP stands for true positive and FN for false negative.
4. F1-score: By taking precision and recall into account, the F1-score offers a fair assessment of a model's performance. This is the harmonic mean of recall and precision [59].

$$\text{F1-score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$
 is the formula to compute F1-score.

5. Specificity: This refers to the percentage of accurate negative predictions among all real negatives, which indicates how well the model can detect negative occurrences. Specificity is calculated as $\text{TN} / (\text{TN} + \text{FP})$, where TN stands for true negatives (negatives that are accurately predicted) and FP stands for false positives (positives that are wrongly predicted).
6. AUC: The area under the receiver operating characteristic (ROC) curve is used to generate the AUC metric. Plotting the true positive rate against the false positive rate at different categorization thresholds results in the ROC curve. An algorithm's capacity to discriminate across classes is shown by its AUC value; a higher AUC value corresponds to stronger discrimination [59, 60].

Model optimization, cross-validation, calibration, and kernel techniques

Model optimization

To maximize the performance of the model, we conducted a comprehensive analysis of the hyperparameters using three common tuning approaches: grid search, random search, and Bayesian optimization [61, 62]. For the grid search, we defined a parameter grid that included key hyperparameters such as learning rate, number of estimators, max depth, and min samples split. We systematically tested a range of values for each hyperparameter and performed an exhaustive evaluation of all possible combinations to identify the optimal configuration.

The random search approach involved randomly sampling hyperparameter values from predefined distributions, allowing us to explore a larger portion of the hyperparameter space in a more efficient manner.

In addition, we leveraged Bayesian optimization, which models the objective function (model performance) as a Gaussian process and intelligently selects the next hyperparameter values to test based on the observed performance so far.

After comparing the results from these three optimization techniques, we found that the grid search approach yielded the best overall performance for our specific dataset and problem. Therefore, we selected the grid search as the primary method for hyperparameter tuning.

Cross-validation

To ensure a thorough and reliable evaluation of the model's performance, we employed three common cross-validation techniques: 3-fold, 5-fold, and 10-fold cross-validation [63].

In the 3-fold cross-validation, we divided the training data into three equal-sized partitions, trained the model on two partitions, and evaluated it on the held-out third partition. We repeated this process three times, using a different partition for the test set each time.

The 5-fold and 10-fold cross-validation followed a similar logic but with five and ten partitions, respectively. By using these varying cross-validation approaches, we were able to obtain more stable and reliable performance estimates, as the smaller the number of folds, the higher the variance in the performance metrics.

After analyzing the results from the different cross-validation methods, we determined that the 10-fold cross-validation provided the most accurate and consistent evaluation of the model's performance for our specific dataset.

Model calibration

To improve the model's accuracy and reliability, we conducted model calibration using isotonic regression [64]. Isotonic regression is a non-parametric technique that adjusts the model's output probabilities to better align with the true likelihood of the predicted outcomes.

The calibration process involved fitting an isotonic regression model to the model's output probabilities and the true class labels. This allowed us to transform the original probabilities in a way that minimized the discrepancy between the predicted probabilities and the actual probabilities of the target classes.

By calibrating the model's output probabilities, we were able to enhance the reliability of the model's predictions and ensure that the reported probabilities accurately represented the likelihood of the outcomes.

Kernel approaches for SVM

In addition to the other models explored, we also investigated the impact of different kernel functions on the performance of the Support Vector Machine (SVM) algorithm. We evaluated five common kernel types: linear, polynomial, radial basis function (RBF), sigmoid, and Gaussian [65].

By comparing the performance of SVM models with these various kernel types, we were able to identify the optimal choice for our specific problem and dataset. The polynomial kernel emerged as the top performer, allowing for non-linear decision boundaries by mapping the input features to a higher-dimensional space using a polynomial function [66].

To quantify the performance of the different kernel functions, we analyzed the accuracy and ROC values, as well as other relevant metrics. The linear kernel achieved an accuracy of 71.0% and an ROC value of 0.76. In contrast, the polynomial kernel provided the highest accuracy at 76.0% and an ROC value of 0.83, outperforming the other kernel functions. The RBF kernel had an accuracy of 75.0% and an ROC value of 0.81, while the sigmoid kernel exhibited the lowest performance with an accuracy of 63.0% and an ROC value of 0.61. The Gaussian kernel achieved an accuracy of 69.0% and an ROC value of 0.75.

Based on these comprehensive results, we can conclude that the polynomial kernel was the optimal choice for our problem and dataset, as it demonstrated the highest metrics performance value compared to the other kernel functions evaluated.

Model interpretability

To reveal hidden patterns and correlations in the dataset, we employed association rule mining techniques, which enhanced the interpretability of our model. The well-known Apriori method, which was developed specifically for association rule mining, had to be used for this. By applying this method, we were able to identify frequently occurring item sets and construct important connection rules using metrics like lift, confidence, and support.

By measuring the strength of correlations between variables, the lift measure was able to demonstrate the effect of one variable on the occurrence of another. Confidence proved that association rules were true by showing the frequency with which the consequent variable arose when the antecedent variable was present. The support measure additionally determined the frequency of occurrence of specific item sets in order to corroborate the association findings. By including support, we were able to gain a comprehensive understanding of the relationships and patterns seen in the dataset, which enhanced the interpretability of our model [67–69].

Table 1 The socio-demographic and economic characteristics of study participants in Sub-saharan Africa based on recent DHS (2015–2024)

Variable	Category	Frequency (n)	Percent (%)
Residence	Urban	53,273	38.15
	Rural	86,386	61.85
Age	15–19	8,686	6.22
	20–29	70,275	50.32
	30–39	49,959	35.77
	40–49	10,739	7.69
Educational level of mother	no education	37,756	27.03
	Primary	49,797	35.66
	Secondary	44,121	31.59
	Higher	7,985	5.72
Sex of household head	Female	32,396	23.20
	Male	107,263	76.80
Religion	Catholic	46,920	33.60
	Methodist	46,092	33.00
	Assembly of God	12,232	8.76
	Universal	8,286	5.93
	Jehovah	11,642	8.34
	Protestant	7,473	5.35
	Islamic	3,581	2.56
	Animist	1,779	1.27
Marital status	Others	1,654	1.18
	Single	11,335	8.12
	Married	118,481	84.84
	Divorced/separated/not living together	9,843	7.05
Wealth index	Poor	53,324	38.18
	Middle	28,561	20.45
	Rich	57,774	41.37
Owned mobile phone	No	59,785	42.81
	Yes	79,874	57.19
Working status	No	49,207	35.23
	Yes	90,452	64.77
Distance to health facility	Big problem/ far from health facility	49,672	35.57
	Not a big problem/ near or optimal	89,987	64.43
Media Exposure	No	53,624	38.40
	Yes	86,035	61.60
Smoke	No	138,483	99.16
	Yes	1,176	0.84
Number of children in the family	Four and below	134,578	96.36
	Above four	5,081	3.64
Region	West Africa	49,641	35.54
	East Africa	71,728	51.36
	Central Africa	15,304	10.96
	South Africa	2,986	2.14

We also employed SHAP (Shapley Additive Explanations) analysis to show how each feature affects the predictive model. SHAP analysis is a useful technique for figuring out the importance and input of specific characteristics in a machine learning model.

Results

Descriptive results of the socio-demographic and economic characteristics

The study conducted a thorough analysis of the descriptive and sociodemographic parameters of 139,659 women who were childbearing. The results showed that the majority of participants—70,275 individuals, or 50.32% of the sample were in the 20–29 age group. Furthermore, 86,386 people, or 61.85% of the poll participants, were rural residents. For more detailed information, please refer to Table 1.

Maternal obstetrics characteristics of the study participants in Sub-saharan Africa

The data showed that the majority of study participants, 91.38% (127,618 individuals), delivered their babies through vaginal delivery. Additionally, 93.04% (129,933 participants) received their delivery services at health institutions. The data also showed that a substantial number, 59.74% (83,426 of women), had never used contraception. This suggests a relatively high rate of non-use of contraceptive methods among the study participants. For more detailed information, please refer to Table 2.

Adverse birth outcome in Sub-Saharan Africa

According to our research of the most current DHS dataset, about 28.59% (95% CI: 28.36, 28.83) of childbearing women in Sub-Saharan Africa had adverse birth outcomes. Our finding also revealed significant variation between each country. Sierra Leone had the lowest prevalence of adverse birth outcomes at 14.08% (95% CI: 13.22 – 14.94%), while Mali had the highest at 45.78% (95% CI: 44.10 – 47.45%). Figure 2 provides a detailed breakdown of the adverse birth outcome rates for each country in Sub-Saharan Africa, sequenced from the highest to the lowest prevalence.

Machine learning analysis of adverse birth outcome

Data balancing

Our analysis of machine learning algorithms using various data balancing techniques is shown in Table 3. The outcome showed that, in comparison to other data balancing strategies utilized in this investigation, SMOTE-ENN consistently outperformed all models evaluated in terms of accuracy. The random forest approach achieved the highest accuracy (87.0%) and AUC (95.0%) notably after data were balanced using SMOTE-ENN. For additional research and optimization, we decide to use

Table 2 Maternal obstetrics characteristics of the study participants in Sub-Saharan Africa based on recent DHS (2015–2024)

Variable	Category	Frequency (n)	Percent (%)
Delivery type	Vaginal	127,618	91.38
	Caesarian	12,041	8.62
Sex of child	Male	71,193	50.98
	Female	68,466	49.02
Parity	Primiparous /1st_time_mothers	26,805	19.19
	Multiparous	75,755	54.24
	Grand multiparous	37,099	26.56
Ever use contraceptive	No	83,426	59.74
	Yes	56,233	40.26
Iron Supplementation	Supplemented	94,377	67.58
	Not Supplemented	45,282	32.42
Type of pregnancy	Single	134,341	96.19
	Twin	5,318	3.81
ANC visit	No visit	3,392	3.18
	Less than three visit (1–3)	64,111	45.91
	Four and above visit	72,156	51.67
Birth interval	Short birth interval	15,908	11.39
	Normal birth interval	103,668	74.23
	Long birth interval	20,083	14.38
Wanted pregnancy	Not wanted	10,175	7.29
	Wanted	129,484	92.71
First ANC check	Early	47,651	34.12
	Late	92,008	65.88
Place of delivery	Home	9,726	6.96
	Health institution	129,933	93.04

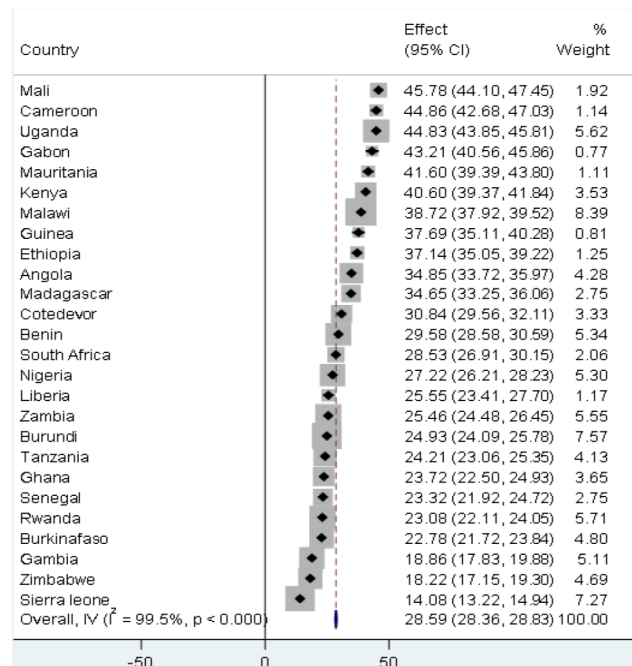


Fig. 2 The prevalence of adverse birth outcome among childbearing women in Sub-Saharan Africa countries using forest tree plot

Table 3 Comparison of imbalanced data handling techniques across each machine learning algorithms

Algorithms	Performance metrics	Unbalanced Data	Under-sampling	Over-sampling	ADASYN	SMOTE	SMOTE-ENN	Near miss
SVM	Accuracy	73.0%	59.0%	60.0%	60.0%	64.0%	76.0%	71.0%
	AUC	0.58	0.63	0.64	0.64	0.70	0.83	0.77
Logistic regression	Accuracy	73.0%	57.0%	57.0%	58.0%	61.0%	70.0%	67.0%
	AUC	0.61	0.60	0.60	0.61	0.65	0.75	0.72
GNB	Accuracy	70.0%	56.0%	56.0%	56.0%	59.0%	62.0%	65.0%
	AUC	0.59	0.60	0.60	0.60	0.64	0.73	0.72
Decision tree	Accuracy	63.0%	54.0%	66.0%	61.0%	66.0%	82.0%	61.0%
	AUC	0.54	0.54	0.67	0.62	0.66	0.80	0.61
Random forest	Accuracy	71.0%	57.0%	70.0%	66.0%	72.0%	87.0%	66.0%
	AUC	0.58	0.60	0.76	0.71	0.79	0.95	0.72
XGBoost	Accuracy	73.0%	59.0%	61.0%	61.0%	67.0%	78.0%	72.0%
	AUC	0.60	0.63	0.66	0.65	0.74	0.86	0.78
KNN	Accuracy	72.0%	56.0%	58.0%	58.0%	62.0%	75.0%	66.0%
	AUC	0.57	0.58	0.62	0.62	0.67	0.82	0.70
AdaBoost	Accuracy	73.0%	59.0%	58.0%	58.0%	64.0%	72.0%	69.0%
	AUC	0.61	0.62	0.62	0.62	0.70	0.78	0.75
FFNN	Accuracy	73.0%	59.0%	60.0%	60.0%	64.0%	76.0%	71.0%
	AUC	0.59	0.63	0.64	0.64	0.70	0.83	0.77
CatBoost	Accuracy	73.0%	60.0%	62.0%	61.0%	67.0%	78.0%	72.0%
	AUC	0.61	0.64	0.66	0.66	0.74	0.86	0.78

SMOTE-ENN. Please see Additional Material 3 for detailed information on the performance metrics of several machine learning algorithms experimenting with different data balancing techniques.

Development and performance comparisons of machine learning models

We carried out various optimization techniques to improve the performance of these machine learning models. We applied random search tuning, Bayesian optimization, and grid search hyperparameter tuning to explore the performance of each algorithm across the different optimization techniques.

Our results showed that grid search optimization produced performance metrics that were consistently higher and more consistent for every algorithm implemented in our study. While Bayesian optimization demonstrated higher performance metrics for a few of the algorithms, specifically KNN and CatBoost, it lacked the overall consistency observed with grid search tuning.

Additionally, the random search tuning technique showed higher performance for the CatBoost algorithm. However, we ultimately chose grid search tuning as the final model optimization technique for the ten machine learning algorithms in our study. This decision was based on grid search’s ability to consistently improve the performance metrics across all the algorithms. Random forest was the top-performing algorithm in all three of the tuning techniques.

For a detailed comparison of the machine learning algorithm performance under the random search, Bayesian optimizations, and grid search tuning approaches, please refer to Table 4. The table presents the accuracy and AUC metrics for each algorithm, after applying ten-fold cross-validation and calibration during hyperparameter optimization steps.

After extensive experimentation and evaluation, we identified the random forest model as the top performer machine learning model for our dataset, achieving outstanding performance with an AUC score of 0.95,

accuracy=0.88, recall=0.94, specificity=0.87, precision=0.88, and F1-score=0.91. For a detailed analysis of the performance metrics, please refer to Fig. 3 (ROC curve) and Fig. 4 (overall metric performance) in our study.

Model interpretability

Association rule mining

The principal predictor variables associated with adverse birth outcomes were determined using association rule mining. The main factors linked to a high risk of adverse birth outcomes according to our exploration with association rule mining were found to be home deliveries, lack of prenatal iron supplementation, fewer than four ANC visits, short and long delivery intervals, unwanted pregnancy, primiparous (first-time) moms, and geographic location in the West African region.

The top 5 association rules and their corresponding statistics were:

Rule 1 If mothers do not have iron supplementation, the pregnancy type is unwanted, and the delivery occurs at home, the probability of an adverse birth outcome is 94.9% (confidence=0.949 and lift=2.96).

Rule 2 If mothers have fewer than four prenatal visits, the pregnancy is unwanted, and the delivery occurs at home, the probability of an adverse birth outcome is 92.5% (confidence=0.925 and lift=2.88).

Rule 3 If mothers have a short birth interval, the pregnancy is unwanted, and the delivery occurs at home, the probability of adverse birth outcomes is 91.7% (confidence=0.917 and lift=2.86).

Rule 4 If mothers are primiparous, the pregnancy is unwanted, and their geographic location is the West Africa region, the probability of adverse birth outcome is 90.8% (confidence=0.908 and lift=2.83).

Table 4 Accuracy and AUC metrics for selected machine learning algorithms after data balancing and hyperparameter optimization

Algorithms	Grid search		Random search		Baysian optimization	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
SVM	84.0%	0.89	81.0%	0.86	83.0%	0.87
Logistic regression	72.0%	0.77	72.0%	0.76	72.0%	0.76
GNB	67.0%	0.72	64.0%	0.72	66.0%	0.72
Decision tree	82.0%	0.80	79.0%	0.80	75.0%	0.79
Random forest	88.0%	0.95	87.0%	0.94	84.0%	0.92
XGBoost	82.0%	0.88	82.0%	0.88	78.0%	0.86
KNN	86.0%	0.90	81.0%	0.84	86.0%	0.91
AdaBoost	72.0%	0.77	70.0%	0.76	72.0%	0.77
FFNN	81.0%	0.84	79.0%	0.83	80.0%	0.84
CatBoost	80.0%	0.86	86.0%	0.92	82.0%	0.88

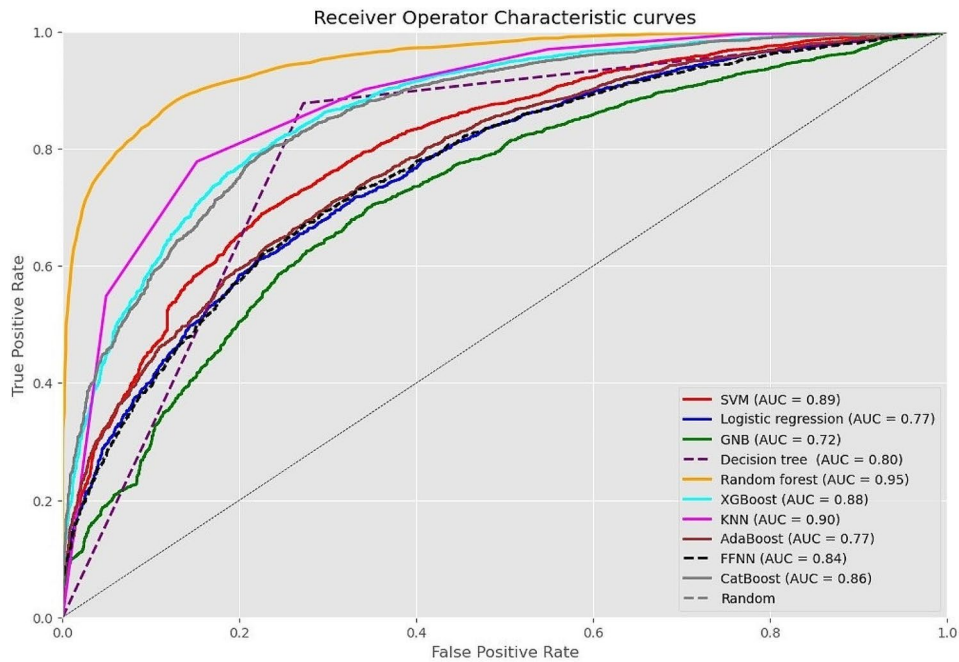


Fig. 3 ROC curve trained with balanced data and optimized model using grid search tuning

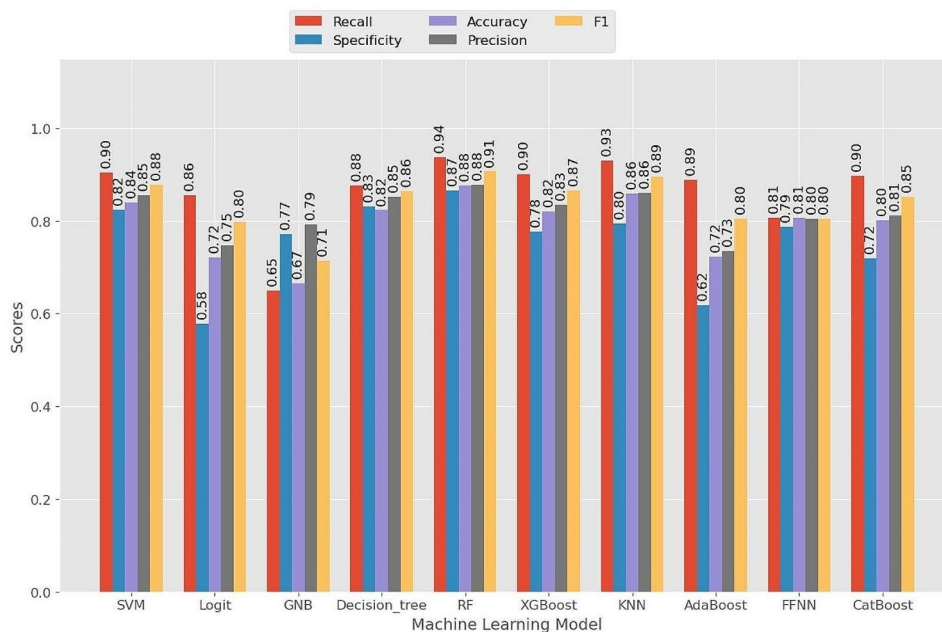


Fig. 4 Comprehensive metrics performance of employed machine learning algorithms using balanced data and optimized model using grid search tuning

Rule 5 If mothers have a long birth interval, the pregnancy is unwanted, and the delivery occurs at home, the probability of adverse birth outcomes is 90.1% (confidence=0.901 and lift=2.81).

SHAP feature impact on model prediction

As depicted in Fig. 5, the top influential factors for the model predictions were region, ANC visit, and parity,

place of delivery, mother education, wanted pregnancy, and iron supplementation.

Discussion

To predict adverse birth outcomes in Sub-Saharan Africa, our research study focused on applying advanced machine learning techniques. The results of our investigation showed significant variables linked to adverse

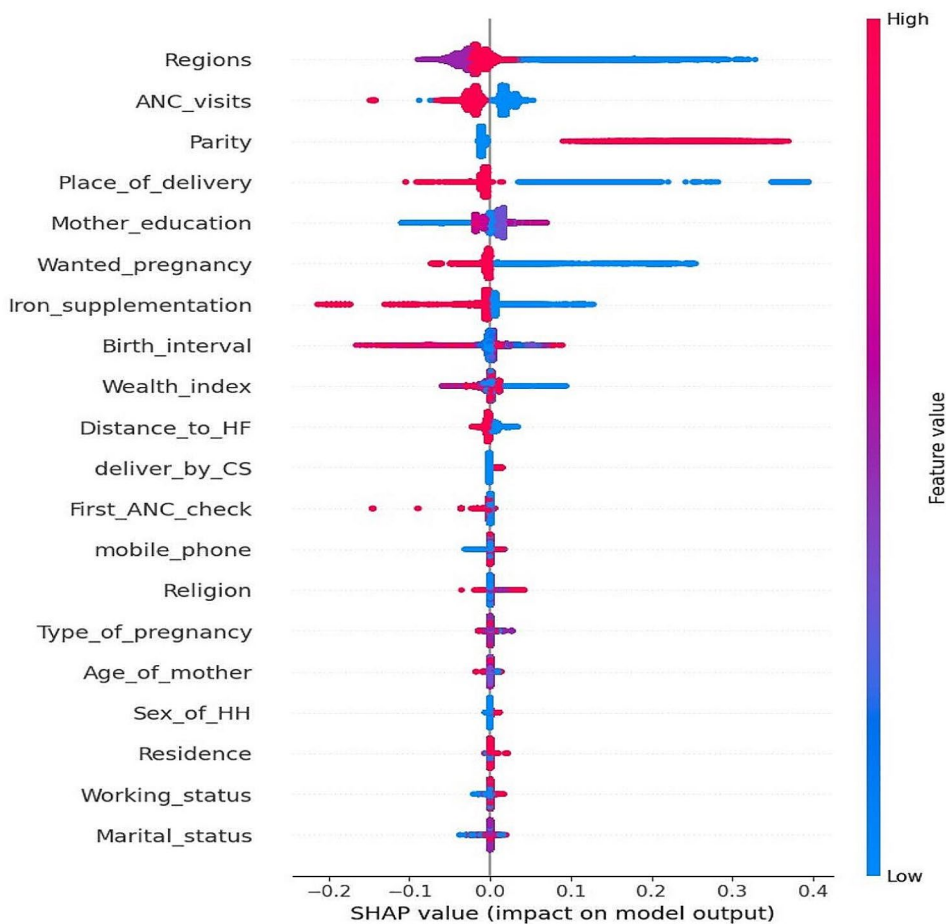


Fig. 5 SHAP feature impact on model prediction

birth outcomes and demonstrated the effectiveness of ten advanced machine learning algorithms in this regard.

According to our findings based on the analysis of the most recent DHS dataset, approximately 28.59% (95% CI: 28.36, 28.83) of childbearing women in sub-Saharan Africa experienced adverse birth outcomes. This prevalence is slightly lower than the previously reported rate for the region [16]. The difference in adverse birth outcome rates might be attributable to the varying periods in which the individual country surveys were conducted.

Regarding the employed algorithm, the findings indicated that the random forest approach was the best-performing algorithm. This finding is consistent with other studies [70–74], suggesting that the random forest method is a valuable tool for predicting adverse birth outcomes. The random forest performs well due to its ability to handle complex and nonlinear relationships within the data [75]. This finding implies that advanced machine learning techniques, specifically the random forest algorithm, hold immense potential as powerful tools to tackle this critical public health challenge of adverse birth outcomes in Sub-Saharan Africa.

Our association rule mining approach discovered multiple factors associated with poor birth outcomes in Sub-Saharan Africa. Inadequate iron supplementation was found to be a major risk factor, indicating the need for iron to encourage healthy fetal development and prevent issues. This conclusion underscores the need for therapies and strategies to ensure adequate iron supplementation for pregnant mothers in the area, in line with previous research findings [20, 24, 31].

Unwanted pregnancy was another important factor that increased the likelihood of poor outcomes at birth. The justification for this could be due to unexpected pregnancies presenting women with a variety of challenges, such as psychological stress, inadequate prenatal care, and a delayed antenatal service start date, all of which can have an adverse effect. The finding is consistent with previous research findings [22].

Access to skilled delivery care is essential for a safe delivery experience, as evidenced by the strong association between home delivery and adverse birth outcomes in our data. As other researchers have already discovered [3, 20, 29], this could be because prompt interventions and emergency obstetric care delivered at

healthcare facilities can reduce the risk of complications and improve outcomes for the mother and newborn.

Our finding implied that inadequate ANC visits, particularly less than the recommended four visits, were associated with an increased risk of adverse birth outcomes. This result is in line with previous research done in other parts of the world [16, 18, 23, 25, 26, 30, 32]. The justification for this could be that inadequate ANC visits increase the likelihood of adverse birth outcomes and missed opportunities for early intervention. The finding implies that ANC visits should be improved if pregnant women are to be encouraged to attend the prescribed number of visits.

Our association rule mining results showed that both short and long birth intervals increase the likelihood of adverse birth outcomes. This finding is supported by previous research elsewhere in the world [3, 21, 30, 33, 76]. This could be because long birth intervals are linked to more challenging pregnancies and older mothers, which could account for the observed findings. The mother's body and the growth of the fetus may also be adversely affected by lengthy waits between pregnancies. A short birth interval, however, might not provide the mother adequate time to heal and restore her nutrition stores. The finding implies pregnant women's and their children's health may be improved by encouraging optimal birth intervals.

According to our findings, one possible factor that raised the risk of adverse birth outcomes was being a primipara mother. This finding is consistent with previous research findings [16–19, 22, 27, 29]. The possible justification for this is that first-time mothers can encounter particular difficulties since they lack experience and understanding about pregnancy and childbirth. The finding implies that educating and providing targeted interventions to first-time mothers can help reduce the risk of adverse birth outcomes.

Within Sub-Saharan Africa, our data showed considerable regional differences in adverse birth outcomes. In particular, we discovered that, in comparison to other regions in Sub-Saharan Africa, living in the West Africa region increased the likelihood of adverse birth outcomes. For this study, the West Africa region was defined to include countries such as Burkina Faso, Ivory Coast, Gambia, Ghana, Guinea, Liberia, Senegal, Sierra Leone, Nigeria, Zambia, Zimbabwe, Benin, Mali, and Mauritania.

This finding implies that regional variations in healthcare quality, access, socioeconomic status, and cultural practices throughout Sub-Saharan Africa have a significant impact on the health outcomes of mothers and children. Strengthening this finding, our further analysis implied that Western Africa had the highest rates of unintended pregnancy and late ANC checkups compared

to other sub-regions of Sub-Saharan Africa, which may be one possible justification for it being a high-risk region for adverse birth outcomes. The finding is consistent with previous research elsewhere in the world [19, 22, 34].

Conclusion

The region continues to face persistent adverse birth outcomes, emphasizing the urgent need for increased attention and action. Fortunately, the findings of this study make it abundantly clear that advanced machine learning techniques, specifically the random forest algorithm, hold immense potential as powerful tools for governments and healthcare providers in Sub-Saharan Africa to tackle this critical issue head-on.

Building upon the invaluable insights obtained from these employed innovative techniques, we strongly recommend that certain key groups of women receive more intensive interventions as soon as their pregnancy is confirmed. Specifically, this includes women who are pregnant for the first time, women with a short (less than 24 months) or long (over 59 months) gap between pregnancies, women with unintended pregnancies, and women living in the high-risk western Africa region identified by the study.

These vulnerable women should be immediately provided with iron supplements and scheduled for comprehensive prenatal care. We should also strongly encourage them to deliver their babies in a medical facility or with the assistance of a qualified health professional, even if they had originally planned a home birth.

Strengths and limitations of the study

This study had several key strengths. First and foremost, the researchers conducted a very thorough analysis, testing out ten different advanced machine learning algorithms and carefully tuning the parameters to get the best performance from each one. This rigorous approach allowed identifying the most suitable algorithms for predicting adverse birth outcomes. Additionally, the study provided valuable insights, highlighting the significant factors that contribute to these adverse birth outcomes. The researchers also used innovative machine learning techniques to uncover the key driving indicators.

However, there were limitations to take into account, though. First, the study was dependent on secondary data such as DHS surveys, which might have certain drawbacks and might not have included all relevant variables. As a result, the authors were limited in their ability to fully explore potential internal correlations among the identified risk factors.

The secondary nature of the data also constrained the authors' capacity to investigate the complex causal pathways underlying the observed associations. Future studies utilizing primary data collection would be valuable to

further elucidate these intricate relationships and provide a more comprehensive understanding of the determinants of adverse birth outcomes in sub-Saharan Africa.

Abbreviations

ADASYN	Adaptively Generating Minority Data
ANC	Anta natal care
AUC	Area under the Receiver Operating Characteristic Curve
DHS	Demographic Health Survey
SVM	Support Vector Machine
SMOTE	Synthetic Minority Over-Sampling Technique
SMOTE ENN	Synthetic Minority Over-Sampling Technique with Edited Nearest Neighbor
XGBoost	eXtreme Gradient Boosting
KNN	K Nearest Neighbors
ROC	Receiver Operating Characteristic Curve
SDGs	Sustainable Development Goals
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-024-19566-8>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We would like to thank the measure DHS program for providing the dataset.

Author contributions

HSN, SAM, ABZ, and BM made significant contributions to the conception, design, data collection, supervision, investigation, data analysis, interpretation, and write-up of the manuscript. GAT, BTS, TDA, and JS have contributed to extracting data, preprocessing data, developing the proposal, validation, revising the manuscript, preparing figures, analysis, visualization, and interpretation of data as well. Finally, all authors (HSN, SAM, ABZ, BM, GAT, BTS, TDA, and JS) reviewed and approved the final manuscript.

Funding

The author(s) received no specific funding for this work.

Data availability

The data used for this study can be accessed from the DHS website upon request to the appropriate authority (<https://dhsprogram.com/data/available-datasets.cfm>). Additionally, the source code used for this study can be shared upon reasonable request from the correspondence author.

Declarations

Ethical approval

We obtained approval from the measure DHS program to access and utilize their data for our study. The DHS program is a well-known source of publically available demographic and health data, which can be accessed after registration and obtaining permission through their website (URL: <https://dhsprogram.com/data/available-datasets.cfm>). The data provided by the DHS was anonymized, ensuring the confidentiality of the respondents. The data was exclusively shared with the co-researchers registered in this account. The DHS Program confirmed that their data was collected after obtaining verbal consent from each respondent. Written consent was not used due to the majority of respondents being illiterate, and the study being a population-based survey, which made it infeasible to obtain written consent. Furthermore, the DHS Program stated that the use of verbal consent for their data collection was waived by each country's Health and Nutrition Research Institute Review Board and the National Research Ethics Review Committee. The study was done in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Health Informatics, School of Public Health, College of Medicine and Health Sciences, Woldia University, PO Box 400, Woldia, Amhara, Ethiopia

²University of South-Eastern Norway, Post office Box 235, Kongsberg N-3603, Norway

³Department of Pediatrics and Child Health Nursing, School of Nursing, College of Medicine and Health Science, Woldia University, Woldia, Ethiopia

⁴Department of Maternal and Reproductive Health, School of Nursing, College of Medicine and Health Science, Woldia University, Woldia, Ethiopia

⁵School of Public Health, College of Medicine and Health Science, Dilla University, Dilla, Ethiopia

⁶Department of Pediatric and Child Health, School of Medicine, College of Medicine and Health Science, Wollo University, Dessie, Ethiopia

⁷Rural Health Research Institute, Charles Sturt University, Bathurst, New South Wales, NSW 2800, Australia

⁸School of Health Sciences and Social Work, Griffith University, Q 4215, Queensland, Australia

Received: 1 June 2024 / Accepted: 23 July 2024

Published online: 29 July 2024

References

- Ghosh R, Causey K, Burkart K, Wozniak S, Cohen A, Brauer M. Ambient and household PM_{2.5} pollution and adverse perinatal outcomes: a meta-regression and analysis of attributable global burden for 204 countries and territories. *PLoS Med*. 2021;18(9):e1003718.
- Korenromp EL, Rowley J, Alonso M, Miello MB, Wijesooriya NS, Mahiané SG, et al. Global burden of maternal and congenital syphilis and associated adverse birth outcomes—estimates for 2016 and progress since 2012. *PLoS ONE*. 2019;14(2):e0211720.
- Tadese M, Dagne K, Wubetu AD, Abeway S, Bekele A, Misganaw Kebede W, et al. Assessment of the adverse pregnancy outcomes and its associated factors among deliveries at Debre Berhan Comprehensive Specialized Hospital, Northeast Ethiopia. *PLoS ONE*. 2022;17(7):e0271287.
- Blencowe H, Krusevec J, De Onis M, Black RE, An X, Stevens GA, et al. National, regional, and worldwide estimates of low birthweight in 2015, with trends from 2000: a systematic analysis. *Lancet Global Health*. 2019;7(7):e849–60.
- Doherty T, Kinney M. Low birthweight: will new estimates accelerate progress? *Lancet Global Health*. 2019;7(7):e809–10.
- Ohuma EO, Moller A-B, Bradley E, Chakwera S, Hussain-Alkhatieb L, Lewin A, et al. National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis. *Lancet*. 2023;402(10409):1261–71.
- Walani SR. Global burden of preterm birth. *Int J Gynecol Obstet*. 2020;150(1):31–3.
- Hughes MM, Black RE, Katz J. 2500-g low birth weight cutoff: history and implications for future research and policy. *Matern Child Health J*. 2017;21:283–9.
- Okwaraji YB, Krusevec J, Bradley E, Conkle J, Stevens GA, Gatica-Domínguez G, et al. National, regional, and global estimates of low birthweight in 2020, with trends from 2000: a systematic analysis. *Lancet*. 2024;403(10431):1071–80.
- Krishna A, Fink G, Berkman LF, Subramanian S. Short-and long-run associations between birth weight and children's height. *Econ Hum Biology*. 2016;21:156–66.
- Bezie MM, Tesema GA, Seifu BL. Multilevel multinomial regression analysis of factors associated with birth weight in sub-saharan Africa. *Sci Rep*. 2024;14(1):9210.
- Harrison MS, Goldenberg RL, editors. Global burden of prematurity. *Seminars in fetal and neonatal medicine*. Elsevier; 2016.
- Hug L, You D, Blencowe H, Mishra A, Wang Z, Fix MJ, et al. Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: a systematic assessment. *Lancet*. 2021;398(10302):772–85.

14. Organization WH. Health in 2015: from MDGs, millennium development goals to SDGs, sustainable development goals. 2015.
15. Bahl R, Martinez J, Bhandari N, Biloglav Z, Edmond K, Iyengar S et al. Setting research priorities to reduce global mortality from preterm birth and low birth weight by 2015. *J Global Health*. 2012;2(1).
16. Tamirat KS, Sisay MM, Tesema GA, Tessema ZT. Determinants of adverse birth outcome in Sub-saharan Africa: analysis of recent demographic and health surveys. *BMC Public Health*. 2021;21:1–10.
17. Gebremeskel F, Gultie T, Kejela G, Hailu D, Workneh Y. Determinants of adverse birth outcome among mothers who gave birth at hospitals in Gamo Gofa Zone, Southern Ethiopia: a facility based case control study. *Qual Prim Care*. 2017;25(5):259–66.
18. Kassahun EA, Mitku HD, Getu MA. Adverse birth outcomes and its associated factors among women who delivered in North Wollo zone, northeast Ethiopia: a facility based cross-sectional study. *BMC Res Notes*. 2019;12:1–6.
19. Khan N, Mozumdar A, Kaur S. Determinants of low birth weight in India: an investigation from the National Family Health Survey. *Am J Hum Biology*. 2020;32(3):e23355.
20. Zaveri A, Paul P, Saha J, Barman B, Chouhan P. Maternal determinants of low birth weight among Indian children: evidence from the National Family Health Survey-4, 2015-16. *PLoS ONE*. 2020;15(12):e0244562.
21. Wulandari F, Mahmudiono T, Rifqi MA, Helmyati S, Dewi M, Yuniar CT. Maternal characteristics and socio-economic factors as determinants of low birth weight in Indonesia: analysis of 2017 Indonesian demographic and Health Survey (IDHS). *Int J Environ Res Public Health*. 2022;19(21):13892.
22. Alamneh TS, Teshale AB, Worku MG, Tessema ZT, Yeshaw Y, Tesema GA, et al. Preterm birth and its associated factors among reproductive aged women in sub-saharan Africa: evidence from the recent demographic and health surveys of sub-saharan African countries. *BMC Pregnancy Childbirth*. 2021;21:1–11.
23. Safitri HO, Fauziningtyas R, Indarwati R, Efendi F, McKenna L. Determinant factors of low birth weight in Indonesia: findings from the 2017 Indonesian demographic and health survey. *J Pediatr Nurs*. 2022;63:e102–6.
24. Ahammed B, Maniruzzaman M, Ferdausi F, Abedin MM, Hossain MT. Socio-economic and demographic factors associated with low birth weight in Nepal: data from 2016 Nepal demographic and health survey. *Asian J Social Health Behav*. 2020;3(4):158–65.
25. Uwimana G, Elhoumed M, Gebremedhin MA, Azalati MM, Nan L, Zeng L. Association between quality antenatal care and low birth weight in Rwanda: a cross-sectional study design using the Rwanda demographic and health surveys data. *BMC Health Serv Res*. 2023;23(1):558.
26. Tsegaye B, Kassa A. Prevalence of adverse birth outcome and associated factors among women who delivered in Hawassa town governmental health institutions, south Ethiopia, in 2017. *Reproductive Health*. 2018;15:1–10.
27. Lin L, Lu C, Chen W, Li C, Guo YY. Parity and the risks of adverse birth outcomes: a retrospective study among Chinese. *BMC Pregnancy Childbirth*. 2021;21:1–11.
28. Mombo-Ngoma G, Mackanga JR, González R, Ouedraogo S, Kakolwa MA, Manego RZ, et al. Young adolescent girls are at high risk for adverse pregnancy outcomes in sub-saharan Africa: an observational multicountry study. *BMJ open*. 2016;6(6):e011783.
29. Khan JR, Islam MM, Awan N, Muurlink O. Analysis of low birth weight and its co-variants in Bangladesh based on a sub-sample from nationally representative survey. *BMC Pediatr*. 2018;18:1–9.
30. Hailemichael HT, Debelew GT, Alema HB, Weldu MG, Migsina KH. Determinants of adverse birth outcome in Tigray region, North Ethiopia: hospital-based case-control study. *BMC Pediatr*. 2020;20:1–9.
31. Abadiga M, Mosisa G, Tsegaye R, Oluma A, Abdisa E, Bekele T. Determinants of adverse birth outcomes among women delivered in public hospitals of Ethiopia, 2020. *Archives Public Health*. 2022;80(1):12.
32. Cherie N, Mebratu A. Adverse birth outcomes and associated factors among delivered mothers in dessie referral hospital. *North East Ethiopia*. 2018:1–6.
33. Tesema GA, Teshale AB, Yeshaw Y, Angaw DA, Molla AL. Assessing the effects of duration of birth interval on adverse pregnancy outcomes in sub-saharan Africa: a propensity score-matched analysis. *BMJ open*. 2023;13(4):e062149.
34. Lin K, Chern S, Sun J. Mapping the quality of prenatal and postnatal care and demographic differences on child mortality in 26 low to middle-income countries. *World J Pediatr*. 2023;19(9):835–50.
35. Yaya S, Uthman OA, Ekholuenetale M, Bishwajit G, Adjiwanou V. Effects of birth spacing on adverse childhood health outcomes: evidence from 34 countries in sub-saharan Africa. *J Maternal-Fetal Neonatal Med*. 2020;33(20):3501–8.
36. Guide to DHS Statistics DHS-7 (version 2) [Internet]. USAID. 2020 [cited November 24, 2023]. <https://dhsprogram.com/data/Guide-to-DHS-Statistics/index.cfm>.
37. Khan MN, Khanam SJ. The effectiveness of WHO's interpregnancy interval advice. *Lancet Global Health*. 2023;11(10):e1476–7.
38. WHO. WHO recommendations on antenatal care for a positive pregnancy experience: WHO. 2016 [<https://www.who.int/publications/i/item/9789241549912>].
39. Kadhim AI. An evaluation of preprocessing techniques for text classification. *Int J Comput Sci Inform Secur (IJCSIS)*. 2018;16(6):22–32.
40. Warnett SJ, Zdun U. Architectural design decisions for the machine learning workflow. *Computer*. 2022;55(3):40–51.
41. Aguinis H, Gottfredson RK, Joo H. Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Res Methods*. 2013;16(2):270–301.
42. Pollet TV, van der Meij L. To remove or not to remove: the impact of outlier handling on significance testing in testosterone data. *Adapt Hum Behav Physiol*. 2017;3:43–60.
43. Pujianto U, Wibawa AP, Akbar MI, editors. K-nearest neighbor (k-NN) based missing data imputation. 2019 5th International Conference on Science in Information Technology (ICSITech); 2019: IEEE.
44. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inf Decis Mak*. 2016;16:197–208.
45. Liu X, Lei S, Wei Q, Wang Y, Liang H, Chen L. Machine learning-based correlation study between perioperative immunonutritional index and postoperative anastomotic leakage in patients with gastric cancer. *Int J Med Sci*. 2022;19(7):1173.
46. Anand H, Vinodchandra S, editors. Applying correlation threshold on Apriori algorithm. 2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN); 2013: IEEE.
47. Zheng A, Casari A. Feature engineering for machine learning: principles and techniques for data scientists. O'Reilly Media, Inc.; 2018.
48. Al-Shehari T, Alsowail RA. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. *Entropy*. 2021;23(10):1258.
49. Seger C. An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. 2018.
50. Pargent F, Bischl B, Thomas J. A benchmark experiment on how to encode categorical features in predictive modeling. München: Ludwig-Maximilians-Universität München; 2019.
51. Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet*. 2018;19:1–6.
52. Hasan M, Rabbi MF, Sultan MN, Nitu AM, Uddin MP. A novel data balancing technique via resampling majority and minority classes toward effective classification. *TELKOMNIKA (Telecommunication Comput Electron Control)*. 2023;21(6):1308–16.
53. Domingues I, Amorim JP, Abreu PH, Duarte H, Santos J, editors. Evaluation of oversampling data balancing techniques in the context of ordinal classification. 2018 International Joint Conference on Neural Networks (IJCNN); 2018: IEEE.
54. Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JM, Moni MA. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. *Comput Biol Med*. 2021;136:104672.
55. Dhall D, Kaur R, Juneja M. Machine learning: a review of the algorithms and its applications. *Proceedings of ICRIC*. 2019: Recent Innovations in Computing. 2020:47–63.
56. Austin PC, Tu JV, Ho JE, Levy D, Lee DS. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *J Clin Epidemiol*. 2013;66(4):398–407.
57. Muraina I, editor. Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts. 7th International Mardin Artuklu Scientific Research Conference; 2022.
58. Tan J, Yang J, Wu S, Chen G, Zhao J. A critical look at the current train/test split in machine learning. *arXiv Preprint arXiv:210604525*. 2021.
59. Erickson BJ, Kitamura F. Magician's corner: 9. Performance metrics for machine learning models. *Radiological Society of North America*; 2021. p. e200126.
60. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther*. 2020;51(5):675–87.
61. Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:181112808*. 2018.

62. Yu T, Zhu H. Hyper-parameter optimization: a review of algorithms and applications. *arXiv Preprint arXiv:200305689*. 2020.
63. Xu Y, Goodacre R. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J Anal Test*. 2018;2(3):249–62.
64. Dormann CF. Calibration of probability predictions from machine-learning and statistical models. *Glob Ecol Biogeogr*. 2020;29(4):760–5.
65. Roman I, Santana R, Mendiburu A, Lozano JA. In-depth analysis of SVM kernel learning and its components. *Neural Comput Appl*. 2021;33(12):6575–94.
66. Vinge R, McKelvey T, editors. Understanding support vector machines with polynomial kernels. 2019 27th European signal processing conference (EUSIPCO); 2019: IEEE.
67. Al-Maolegi M, Arkok B. An improved Apriori algorithm for association rules. *arXiv Preprint arXiv:14033948*. 2014.
68. Hussein N, Alashqur A, Sowan B. Using the interestingness measure lift to generate association rules. *J Adv Comput Sci Technol*. 2015;4(1):156.
69. Bao F, Mao L, Zhu Y, Xiao C, Xu C. An improved evaluation methodology for mining association rules. *Axioms*. 2021;11(1):17.
70. Bekele WT. Machine learning algorithms for predicting low birth weight in Ethiopia. *BMC Med Inf Decis Mak*. 2022;22(1):232.
71. Ahmadi P, Alavimajd H, Khodakarim S, Tapak L, Kariman N, Amini P, et al. Prediction of low birth weight using Random Forest: a comparison with logistic regression. *Archives Adv Biosci*. 2017;8(3):36–43.
72. Faruk A, Cahyono ES, Eliyati N, Arifeni I. Prediction and classification of low birth weight data using machine learning techniques. *Indonesian J Sci Technol*. 2018;3(1):18–28.
73. Zahirzada A, Lavangnananda K, editors. Implementing predictive model for low birth weight in Afghanistan. 2021 13th International Conference on Knowledge and Smart Technology (KST); 2021: IEEE.
74. Cho H, Lee EH, Lee K-S, Heo JS. Machine learning-based risk factor analysis of adverse birth outcomes in very low birth weight infants. *Sci Rep*. 2022;12(1):12119.
75. Rigatti SJ. Random forest. *J Insur Med*. 2017;47(1):31–9.
76. Agrawal S, Chaudhary M, Das V, Agarwal A, Pandey A, Kumar N, et al. Association of long and short interpregnancy intervals with maternal outcomes. *J Family Med Prim Care*. 2022;11(6):2917–22.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.