RESEARCH



Exploring machine learning algorithms to predict short birth intervals and identify its determinants among reproductive-age women in East Africa



Tirualem Zeleke Yehuala^{1*}, Bezawit Melak Fente² and Sisay Maru Wubante¹

Abstract

Background The occurrence of short birth intervals among reproductive-age women in East Africa is a critical public health issue, contributing to maternal and child health risks. Identifying the key factors that predict short birth intervals can help design targeted interventions to reduce these risks. Hence, this study aimed to predict short birth intervals and identify their determinants using supervised machine learning models.

Method This study employs machine learning algorithms to predict short birth intervals among reproductive-age women in East Africa, using a dataset from Demographic and Health Surveys. The dataset undergoes preprocessing steps to handle missing values, encode categorical variables, perform feature selection, and integrate data and normalize numerical features. Four machine learning models, including logistic regression, decision trees, random forests, and some machine learning models, including logistic regression, decision trees, random forests, are trained and evaluated to predict short birth intervals. Model performance is assessed using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC used to ensure reliable results.

Result The machine learning models identified several key factors that significantly predict short birth intervals among reproductive-age women in East Africa. The Random Forest models demonstrated the highest accuracy (79.4%), precision (79.0%), F-score (84.0%), ROC curve (83.8%), and recall (91.0%), with feature importance analysis highlighting maternal age, educational status, parity, use of family planning, and access to healthcare as the most influential predictors. The findings underscore the importance of targeted interventions addressing healthcare access and family planning to reduce the risks associated with short birth intervals in East African countries.

Conclusion The study demonstrates that machine learning models can effectively identify key predictors of short birth intervals among reproductive-age women in East Africa, providing valuable insights for designing targeted public health interventions to improve maternal and child health outcomes in East Africa.

Keywords Prediction, Short birth interval, Supervised machine learning, East Africa

*Correspondence: Tirualem Zeleke Yehuala sarazeleke3@gmail.com



¹Department Health informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia ²Department of General Midwifery, School of Midwifery, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

© The Author(s) 2025. **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://creative.commons.org/licenses/by-nc-nd/4.0/.

Introduction

Birth interval is an important part of family planning (FP) and fertility control [1]. According to the World Health Organization (WHO), short birth intervals are defined as less than 33 months between two consecutive live births [2, 3]. This issue has received special attention in public health and family planning due to its impact on fertility, maternal, and child health [4]. The practice of short birth spacing varies globally. It increases the risk of adverse effects on maternal and child health outcomes, including insufficient folate levels, inadequate breastfeeding for the child, infections in the cervix, competition between siblings, insufficient uterine repair, and aberrant endometrial blood vessel remodeling [5, 6]. In addition, closely spaced births result in high fertility, which raises the demand for resources from women and inhibits economic development efforts by restricting women's participation [4, 7].

Annually, there are more than 2.5 million perinatal deaths worldwide, 95% of which occur in developing countries, and short intervals between births are directly related to a higher risk of unfavorable perinatal outcomes [8]. Similar to this, numerous poor perinatal outcomes have been attributed to short birth intervals in sub-Saharan African countries [9, 10]. The highest incidence of unfavorable pregnancy outcomes was perinatal events, which include stillbirth, low birth weight, preterm birth, and small for gestational age, and these consequences are widespread in developing countries (9, 5).

Almost 200 million reproductive-age women in lowand middle-income countries want to have space or limit their pregnancies, and access to family planning is especially limited in East African countries [5]. East African countries have the highest rate of short birth gaps (Uganda: 13.4% [11], Ethiopia: 58.5% [12], Tanzania: 48.4% [2], responsible for the huge burden of child and maternal mortality [13–15]. An estimated 1.6 million deaths of under-five mortality would be averted annually if all births spaced 24 months and above [16].

Previous studies found age, women's health care decision-making autonomy, sex of household head, media exposure, maternal education, household wealth status, husband education, and health care access were the most significant predictors of short birth interval [5, 6, 10–13, 17, 18].

Even though appropriate birth spacing is critical for the health of the mother and newborn, family planning is insufficiently utilized and not widely recognized in East African countries [19, 20]. Previous studies have focused on identifying the primary risk factors that contribute to short birth intervals and the effect of short birth intervals on maternal and neonatal health by utilizing logistic regression and other conventional statistical methods to identify and quantify predictors of birth intervals [16, 21–23].

Traditional statistical methods often struggle with high-dimensional data, non-linear relationships, they often fall short in modeling complex, nonlinear interactions between predictors and outcomes and complex interactions due to their reliance on predefined assumptions linearity, normality. These methods may also suffer from overfitting and limited flexibility in capturing intricate patterns. In contrast, machine learning (ML) overcomes these limitations by automatically modeling non-linear relationships, handling large and complex datasets, improving prediction accuracy, and performing efficient feature selection. This study aims to address these gaps by leveraging ML techniques to enhance and develop model to predict shot birth interval, offering a more robust approach than traditional statistical methods [24]. ML algorithms are typically designed to make accurate predictions by learning from data, rather than relying on prior assumptions. By utilizing machine learning algorithms, this study uncovers hidden patterns and interactions within the data, leading to model building and a deeper understanding of the key determinants of birth intervals. The data-driven insights can inform evidence-based decision-making and resource allocation for treatment-related services and policies. Hence, this study aimed to predict short birth intervals and identify their predictors using a supervised machine-learning model.

Methods

Population and eligibility criteria

All women between the ages of 15 and 49 who were of reproductive age served as the study's source population. The study population consisted of women who had at least two consecutive live births over the five years prior to the survey.

Data source and sampling procedures

In this study, we used the most recent Demographic and Health Surveys (DHSs) dataset from 11 East African countries (Burundi, Ethiopia, Comoros, Uganda, Tanzania, Mozambique, Madagascar, Zimbabwe, Kenya, Zambia, and Malawi) as the source of the data for this study [25]. We accessed this data from the MEASURE DHS program by requesting the official database, http://www .dhsprogram.com to obtain the data.

Among East African countries, only 11 countries were included in this study due to the absence of outcome variables, and it was not conducted DHS recently in the remaining countries. DHS data is a nationally representative household survey that is collected periodically in various groups. To select study participants, a two-stage stratified cluster sampling technique was used. In the first stage, a stratified sample by geographic region and urban/rural areas within each region that entirely covers the target population of East Africa was used. In the second stage, households were selected using systematic random sampling in the selected EAs. In each selected household, mothers were interviewed with an individual questionnaire. After applying data preparation, 100,246 weighted samples were included in this study.

Study variables and measurements

Dependent variables: In this study, the dependent variable was birth spacing, categorized as short versus optimal. The birth interval was defined as the time between the preceding birth and the index birth, measured in months from the mother's previous childbirth to the current birth. Therefore, it was dichotomized as 0 = "yes" (having a short birth interval) and 1 = "no" (having an optimal birth interval). Birth intervals were then categorized as short birth intervals if they occurred within 33 months of the preceding birth, and optimal birth intervals if they occurred at 33 months or later.

Independent variables In this study, we used the fast recursive feature selection method to identify the independent features.

Data analysis procedure

In this study, data processing, data analysis, and model building were carried out using Python, with libraries like Pandas for cleaning, encoding categorical variables, and normalizing numerical features. After splitting the data into training and testing sets using train_test_split, exploratory data analysis was performed with visualizations from matplotlib and seaborn to understand feature distributions and relationships. Models were built using scikit-learn, where algorithms such as Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and Naive Bayes (NB) were trained on the data. Hyperparameter tuning and model evaluation were carried out using techniques like grid search, and performance was assessed through metrics such as accuracy, precision, recall, and F1-score. Finally, we developed a predictive model that predicts the short birth interval and identifies its determinants, as shown in Fig. 1.

Data preprocessing

In this study, we employed key data preprocessing techniques such as data cleaning, transformation, integration, feature selection, and discretization, which include handling missing values, encoding categorical variables, and normalizing or scaling features. These steps ensure that



Fig. 1 Overview of data pre-processing and model development to predict short birth interval

data is properly prepared for building accurate and efficient models.

Data cleaning

Data cleaning typically includes handling missing values through imputation or removal, removing outliers, and removing noise-missing values [26]. Additionally, duplicate records were eliminated, and data was standardized or normalized to ensure uniformity, improving the model's ability to learn effectively and accurately. Among the total number of data sets, 3.6% are missing data points. Features such as distance of health facility, health insurance, unmet need, and husband education status have missing values. We used mode imputation for categorical variables to address these missing values.

Data transformation

In machine learning, data transformation refers to the process of changing or modifying raw data into a format that is more suitable for modeling [27]. In this study, we employed data normalization to transform features to a common scale, typically between 0 and 1. Additionally, we used one-hot encoding to transform categorical data into numerical.

Data discretization

Data discretization is a technique used to transform continuous data into discrete categories or intervals. It can help simplify the model, improve interpretability, and enhance performance in certain cases. In this study, we used equal-width binning, where the range of continuous values is divided into a set number of equal-width intervals. The maternal age was discretized as 15–24, 25–34, and 35–49 years based on DHS guidelines [23].

Data integration

Data integration involves combining data from different sources or formats into a cohesive dataset suitable for analysis. In this study, datasets from 11 different sources in East African countries' DHS databases were merged based on common determinants. This approach allows for the creation of a comprehensive dataset.

Feature selection

Feature selection is crucial, and machine-learning models such as RF, DT, LR, and NB can be used to rank features by their importance in predicting birth intervals. This helps in selecting the most impactful variables. In this study, we utilized Recursive Feature Elimination (RFE), a method in which features are recursively removed from the model based on their importance until the optimal subset of features is identified.

Mode selection

The predicted variable in this study was binary classification since the birth interval was divided into two "optimal" or "short." For model building, four classifiers Random Forest, Decision Tree, Logistic Regression, and Naive Bayes—were used. The algorithms were chosen by previous research that used machine-learning methods to predict binary tasks [28, 29]. Random forests and decision trees are powerful for binary prediction due to their ability to capture complex, non-linear relationships in the data. Logistic regression and naive Bayes, on the other hand, are simpler models that provide probabilistic interpretations and are effective when the relationships between variables are linear or when assumptions like independence hold.

Data splitting

Data splitting is a vital part of machine learning to accurately assess a model's performance. The goal is to partition the dataset into separate subsets for training and testing, allowing the model to be evaluated on its ability to generalize to new, unseen data and to reduce the risk of overfitting to the training set. In this study, we used the straightforward holdout method, designating 80% (800,197 cases) of the data for training the model and 20% (200,049 cases) for testing its performance.

Model evaluation

The performance of the trained model was evaluated using the testing dataset. Then, the performance of the trained models was evaluated using the test set based on the criteria of accuracy score, ROC curve, precision (P), recall (R), and F-measure as follows: The confusion matrix is a matrix of N * N, where N is the number of predicted classes, and it displays the number of correct and incorrect predictions made by the classification model relative to the target value [30]. Subsequently, the test set was used to assess the trained models' performance using the accuracy score, AUCROC curve, precision (P), recall (R), and F-measure criteria.

$$Precision = (TP) / (TP + FP)$$
(1)

$$Recall = (TP) / (TP + FN)$$
(2)

$$F - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \quad (3)$$

$$Accuracy = \left(\frac{(TP+TN)}{(TP+TN+FP+FN)}\right) \times 100$$
 (4)

Table 1	Description of the birth interval in East African	
countries, evidence from DHS ($N = 100,246$)		

countries, endence nom	0110 (11 100	(2.10)	
Urban	26,947(27.0)	9109(33.8)	17,865(72.2)
Maternal age			
15–24	13,006(12.9)	7410(57.0)	5596(43.0)
25–34	445,243(44.4)	15,430(34.6)	29,113(63.4)
>=35	42,697(42.5)	16,841(39.4)	25,856(60.6)
Maternal education			
No	24,235(24.0)	11,122(45.9)	13,113(54.2)
Primary	52,662(52.5)	20,945(39.8)	31,717(60.2)
Secondary	20,010(19.9)	6590(32.7)	13,420(67.3)
Higher	3339(3.3)	1024(30.6)	2315(69.4)
Husband education			
No	14,481(14.4)	6,655(46.0)	7,826(54.0)
Primary	58,768(58.6)	23,601(41.1)	35,167 (59.9)
Secondary	22,230 (22.2)	7,852(35.3)	14,378(64.3)
Higher	4,767(4.7)	1,573(33.0)	3,194(67.0)
Wealth status			
Poorest	22,488(22.4)	10,496(46.7)	11,992(53.3)
Poorer	19,522(19.4)	8418(43.1)	11,104(56.9)
Middle	19,049(19.0)	7559(39.7)	11,490(60.3)
Richer	19,346(19.0)	7024(36.3)	12,322(63.7)
Richest	19.841 (19.7)	6184(31.1)	13.657(68.9)
Media exposure	., ,		-,,
No	40,304 (40.2)	16,636 (41.2)	23,668 (58.8)
Yes	59,942 (59.8)	23,045 (38.4)	36,897 (61.6)
Sex of household head			
Wife	29,524(29.4)	10,589(35.8)	18,935(64.2)
Husband	70,722(70.6)	29,092(41.1)	41,630(58.9)
Marital status			
Single	2,028 (2.1)	618 (30.4)	1,410 (69.6)
Married	82,373 (82.1)	33,477 (40.6)	48,896 (59.4)
Divorced/widowed/	15,845 (15.8)	5,586 (35.2)	10,259 (64.8)
separated			
Parity			
2-3	43,603(43.4)	16,319(37.4)	27,284(62.6)
4–5	29,414(29.3)	10,907(37.0)	18,507(63.0)
≥б	27,229(27.3)	12,455(45.7)	14,774(54.3)
Unmet need for family pla	nning		
No	29,141(29.0)	11,499(39.4)	17,642(60.6)
Yes	71,105(71.0)	28,182(39.6)	42,923(60.4)
Distance to health facility			
Not a big problem	36,176(36.0)	14,902 (31.1)	2127 (68.9)
Big problem	64,070(64.0)	24,479(38.2)	439,291(61.8)
Women health care decisi	on making aut	onomy	
Respondent alone	20,416(20.0)	7828(38.3)	12,588(51.7)
Jointly with husband/	62,178(62.4)	24,442(39.3)	37,936(60.7)
partner	/		,
Respondent/partner alone	17,652(17.6)	7611(43.1)	10,041(56.7)
Covered by health insurar	ice		
No	93,714(93.3)	37,422(40.0)	56,294(60.0)
Yes	6532(6.7)	2259(34.6)	4273(65.4)

Results

Sociodemographic characteristics of the study participant

100,246 women of reproductive age who had more than one child were included in the study. The median age of women was 30 years (IQR), and approximately 73,272 (73.0%) were rural residents, with 43.7% having a short birth space. Kenya and Malawi accounted for around 18.7% and 15.1% of the women, respectively. 13,006 (12.9%) of the women were between the ages of fifteen and twenty-four, while 57.0% had a short birth interval. 24,235 (24.0%) had no formal education, whereas 45.9% had a short birth interval. Nearly half (46.7%) of the 22,488 women with the poorest wealth level had short childbearing intervals. 43,603 (43.4%) of the total 100,246 reproductive-age women had two to three births, 37.4% had a short birth interval, and 71,105 (71.0%) had an unmet demand for family planning (Table 1).

Class balancing

In order to balance the target features for this study, we applied the Synthetic Minority Oversampling Technique (SMOTE). This technique generates additional synthetic observations from the minority category in order to balance the unequal distribution of the outcome variable. Prior to SMOTE balancing, the optimal birth interval was 60,565 (60.4%), while the short birth interval was 39,681 (39.9%). We obtained a balanced sample of optimal birth interval with counts of label 60,565 (50%) as shown in Fig. 2.

Identify the important determinants

Important features selection and impact of variables by reducing too many or unnecessary features, rank was a method for finding a subset of features. By limiting the number of features, feature selection was crucial for lowering learning costs. This study used recursive feature elimination. The dataset used in this study contained 25 features. To identify the most significant ones, we used the fast version of RFE, which allows for flexibility in controlling the number of features retained and effectively handles correlated features. The selected features included age, parity, country, wealth status, maternal educational status, sex of household head, media exposure, unmet need, husband education status, and others. The features with longer bars were associated with a higher predicted probability of a short birth interval, showing that these factors increase the likelihood of a short birth interval. Were as the feature with short bars these features were associated with a lower predicted probability of a short birth interval, as shown in Fig. 3.



Fig. 2 Before and after applying SMOTE oversampling techniques



Fig. 3 Important features selected from the dataset by recursive feature elimination

Comparing supervised machine-learning algorithm

The goal of this study was to develop a predictive model for short birth intervals and identify the key determinants for evidence-based decision-making. Supervised machine learning algorithms, including RF, DT, LR, and NB, were utilized. The experiments were conducted using consistent testing parameters. Predictive performance was evaluated using four metrics: accuracy, F-measure, recall, and the AUC curve. After comparing the supervised machine learning models, random forest emerged as the top-performing model, achieving an accuracy of 80%, precision of 79%, recall of 91%, an F-measure of 84%, and an AUC of 83.8%. The second-best algorithm was the decision tree, which achieved an accuracy of 75.3%, precision of 76%, recall of 88%, an F-measure of 81%, and an AUC of 82.7%. The least performing algorithm was logistic regression, which attained an accuracy of 63.2%, a recall of 87%, an F-measure of 74%, and an AUC of 63.5%, as shown in Figs. 4; Table 2. We employed an automated approach for hyperparameter tuning by utilizing the GridSearchCV method. GridSearchCV takes in an estimator, a set of hyperparameters to be searched over, and a scoring method, and returns the best set of hyperparameters that maximizes the scoring method. Before using GridSearchCV, the accuracy was 75.1%; after applying hyperparameter tuning with n_estimators = 200, max_depth = 15, and min_samples_split = 5 accuracy was increased by 4.28% from 75.1 to 79.4%.

Model interpretation using SHAP values

In this study, we use SHAP values to explain the predictions made by a machine learning model, such as a



Fig. 4 AUC curve for machine learning model

 Table 2
 Accuracy, precision, recall, and F-measure for the machine learning algorithms

Model	Accuracy	Precision	Recall	F1-Score	AUC
Random forest	79.4	79.0	91.0	84.0	83.8
Decision tree	75.3	76.0	88.0	81.0	82.7
Logistic Regression	63.2	64.4	87.0	74.0	63.5
Naive Basis	70.0	72.0	81.0	76.0	71.9

random forest model, for predicting whether the birth interval is a binary classification: short interval and optimal interval. SHAP values help us interpret how each feature contributes to the individual predictions, as well as the overall model behavior. In this study, we used SHAP values for model interpretation, since features that push the prediction higher compared to the base value have positive SHAP values, while features that push the prediction lower have negative values. Therefore, for the first observation, the combination of the positive contributions (in red) and the negative contributions (in blue) moves the expected value output to the final model output (f(x) = 0.217), classified as a negative class (short interval). The country was Zimbabwe, the parity was 2–3, and the age of the mother was 25–34; the distance of the health facility was fair, women had health care decision-making by their husbands, the wealth status was poor, they had no media exposure, and the marital status was married all features associated with a higher likelihood of short birth intervals in this specific prediction shown in Fig. 5.

Discussion

In this study, we aimed to predict short birth intervals and identify their key determinants among reproductiveage women in East Africa using a supervised machinelearning model. The use of supervised machine learning models, such as random forest, decision tree, logistic regression, and naive Bayes, allowed for a robust and accurate prediction model. For this objective, four supervised machine-learning models were trained on balanced training data using a train-test split. Accuracy, AUC score, precision, recall, and F1 score were used to compare the performance of classifier models. Random forest emerged as the best predictive model, with an accuracy performance of 79.4%, precision of 79.0%, an F1 score of 84%, and AUROC of 84%. To date, no studies have been conducted utilizing supervised ML algorithms to predict short birth intervals and their determinants. Important predictors' are identified by using RFE methods. We found that the top five variables in line with other studies elsewhere age, parity, country, wealth status, and maternal educational status were all the important variables for the prediction of short birth interval. For instance, when the age of women increased with short birth intervals was decreased. It is consistent with studies [31, 32]. This might be because there are more opportunities as



Fig. 5 Waterfall plot displaying prediction positive and negative values for short birth interval

a woman gets older to interact with medical professionals, which could lead to health education initiatives that encourage better attitudes and advancements in family planning [33]. In addition, older mothers are knowledgeable about and supportive of family planning services to have positive outcomes for both the health of the mother and the child. This study found that parity was an important predictor of short birth intervals. Women who had 2-3 had a higher risk of short birth intervals. It was supported by studies [34, 35], this could be women who are grand multipara have less desire for an extra child and therefore they use family planning methods [36]. Women with higher parity may be more likely to experience short birth intervals, especially if they did not have sufficient time to recover physically from the previous pregnancy before becoming pregnant again. Short birth intervals are associated with a higher risk of preterm birth, low birth weight, and developmental delays. According to the results, there were differences in birth interval rates by wealth status, with women who belonged to the poorest household having higher short birth intervals compared to women from the richest household immunization rates than rural residents. It was supported by previous studies [13, 37, 38], this might be because women from poor households have limited access to maternal health care services such as contraceptive use, and access to contraceptive services could contribute to the increased odds of a short birth interval among women belonging to the poorest household [39]. Additionally, women have Lower education is frequently associated with limited knowledge of family planning options, or less autonomy in decision-making around reproductive health. This can result in higher fertility rates and shorter intervals between births, especially if women lack access to family planning services. Likewise, studies have found that maternal education is a major contributor to short birth intervals [5, 13, 25, 35, 38]. The possible explanations might be due to education, which could raise the maternal and child health implications of short birth intervals and the likelihood of using reproductive health care services to prolong inter-birth intervals [40, 41].

Women aged 15–24 are more likely to experience short birth intervals due to higher fertility, limited access to contraception or family planning resources, and not knowing family planning [42]. This could be attributed to the fact that a woman whose age is 15–24 has a high fertility rate as compared to their counterparts [43]. In addition, older mothers may also experience short birth intervals, often due to pressures to conceive before fertility declines or the use of fertility treatments, despite reduced natural fertility. Several studies agreed with this finding [6, 44, 45].

Conclusion and recommendation

This study aimed to predict short birth intervals and identify their determinants among reproductive-aged women in East Africa using a supervised machine learning method. The random forest algorithm was found to be the best-performing machine learning algorithm model for predicting short birth intervals in East Africa. Important features for this prediction included age, parity, country, and wealth status, unmet need for family planning, distance to health facilities, residence, and maternal educational status. To address the issue of short birth intervals, it is recommended that policymakers focus on enhancing maternal education, particularly in rural areas, and increase access to family planning services. Interventions should be culturally sensitive, taking into account the unique social and economic contexts of East African countries. Strengthening healthcare infrastructure, especially maternal and child health services, will also be critical in supporting women's reproductive decisions. Additionally, incorporating machine-learning models into public health strategies could help design targeted interventions and improve long-term monitoring of birth interval trends, ultimately contributing to better maternal and child health outcomes in East Africa. We recommended that policymakers take into account the findings of this research and provide a strategy for the optimization of birth intervals in low-income countries based on the relevant variables that have been identified. Even though a fascinating result was obtained, future works were required by applying alternative types of techniques with different parameters.

Strength and limitations

The use of machine learning algorithms to identify the best features for predicting short birth intervals among reproductive-age women in East Africa offers significant strengths. These models can handle large, complex datasets and identify intricate patterns between factors such as age, parity, wealth status, maternal education, and country. Machine learning techniques like random forests and logistic regression uncovering non-linear relationships and interactions among these features, offering deeper insights than traditional statistical methods. Moreover, the adaptability of machine learning models allows for tailored interventions based on local contexts, and the ability to rank feature importance enables policymakers to prioritize key variables in improving birth spacing practices. However, this study has certain limitations because the DHS data collection is self-reported, which may have introduced some information biases.

Abbreviations

- CI Confidence Interval
- DHS Demography and Health Survey

AUC Area Under the Curve

FRE	Recursive Feature Elimination
IQR4	Interquartile range
ML	Machine Learning
SHAP	SHapley Additive exPlanations
WHO	World Health Organization

Acknowledgements

We are very grateful to MEASURE DHS authority for they have offered us all the necessary data and documents, which we needed for this work. We would also like to extend our heartfelt gratitude to our friends for their constructive comments, critical readings, encouragement, and suggestions, which are important for the accomplishment of this work.

Author contributions

TZY, BM and SMW made conceptualization and writing the original draft. TZ, BM and SMW made significant contributions in the design, data collection, supervision, data curation, investigation, data analysis, and interpretation of the manuscript. Both authors equally contributed to the design of the proposal, validation, and revision of the article, construction of the figures, analysis, visualization, and interpretation of the data. Both authors reviewed, read, and approved the final manuscript.

Funding

Not applicable.

Data availability

The DHS Program repository has the datasets used in this work, which are openly accessible there. Permission to get access to the data was obtained from the measure DHS program online request from http://www.dhsprogram. com after we had sent the research question.

Declarations

Ethical approval and consent to participate

The measure DHS program gave us permission to download and use the data for our research, and this study is based on secondary data analysis. Thus, participant consent and ethical approval are not required for this particular study. Data were initially gathered by "The DHS program" with the informed agreement of the subjects. The DHS program's official database makes the dataset anonymously accessible to the public.

Consent for publication

It does not apply to this study since the study was secondary data collected by a central statistics agency.

Competing interests

The authors declare no competing interests.

Received: 23 October 2023 / Accepted: 29 April 2025 Published online: 09 May 2025

References

- Kassie SY, Ngusie HS, Demsash AW, Alene TD. Spatial distribution of short birth interval and associated factors among reproductive age women in Ethiopia: Spatial and multilevel analysis of 2019 Ethiopian mini demographic and health survey. BMC Pregnancy Childbirth. 2023;23:275.
- Exavery A, Mrema S, Shamte A, Bietsch K, Mosha D, et al. Levels and correlates of non-adherence to WHO recommended inter-birth intervals in Rufiji, Tanzania. BMC Pregnancy Childbirth. 2012;12:1–8.
- Marston C. (2006) Report of a technical consultation on birth spacing, Geneva, 13–15 June 2005.
- Conde-Agudelo A, Rosas-Bermudez A, Castaño F, Norton MH. Effects of birth spacing on maternal, perinatal, infant, and child health: a systematic review of causal mechanisms. Stud Fam Plann. 2012;43:93–114.
- Pimentel J, Ansari U, Omer K, Gidado Y, Baba MC, et al. Factors associated with short birth interval in low-and middle-income countries: a systematic review. BMC Pregnancy Childbirth. 2020;20:1–17.

- De Jonge HC, Azad K, Seward N, Kuddus A, Shaha S, et al. Determinants and consequences of short birth interval in rural Bangladesh: a cross-sectional study. BMC Pregnancy Childbirth. 2014;14:1–7.
- Towriss CA. Birth intervals and reproductive intentions in Eastern Africa: insights from urban fertility transitions. London School of Hygiene & Tropical Medicine; 2014.
- Karra M, Miller R. Assessing the impact of birth spacing on child health trajectories. Harvard Center for Population and Development Studies, Boston University; 2018.
- Ajayi Al, Somefun OD. (2020) Patterns and determinants of short and long birth intervals among women in selected sub-Saharan African countries. Medicine 99.
- 10. Gonçalves SD, Moultrie TA. Short preceding birth intervals and child mortality in Mozambique. Afr J Reprod Health. 2012;16:29–42.
- Aleni M, Mbalinda S, Muhindo R. (2020) Birth intervals and associated factors among women attending young child clinic in Yumbe hospital, Uganda. International journal of reproductive medicine 2020.
- Leka YL, Feleke FW. Prevalence and predictors of short birth interval among married women in Mareka district, South Ethiopia. EC Gynecol. 2022;11:20–30.
- Alvarez JL, Gil R, Hernández V, Gil A. Factors associated with maternal mortality in Sub-Saharan Africa: an ecological study. BMC Public Health. 2009;9:1–8.
- Muldoon KA, Galway LP, Nakajima M, Kanters S, Hogg RS, et al. Health system determinants of infant, child and maternal mortality: A cross-sectional study of UN member countries. Globalization Health. 2011;7:1–10.
- 15. Rutaremwa G. (2012) Under-five mortality differentials in urban East Africa: a study of three capital cities. Afr Popul Stud 26.
- Rutstein SO. Effects of preceding birth intervals on neonatal, infant and under-five years mortality and nutritional status in developing countries: evidence from the demographic and health surveys. Int J Gynecol Obstet. 2005;89:S7–24.
- Luo Z-C, Wilkins R, Kramer MS. Effect of neighbourhood income and maternal education on birth outcomes: a population-based study. CMAJ. 2006;174:1415–20.
- 18. Whitworth A, Stephenson R. Birth spacing, sibling rivalry and child mortality in India. Soc Sci Med. 2002;55:2107–19.
- Magadi MA, Agwanda AO, Obare FO. A comparative analysis of the use of maternal health services between teenagers and older mothers in sub-Saharan Africa: evidence from demographic and health surveys (DHS). Soc Sci Med. 2007;64:1311–25.
- Odhiambo JN, Sartorius B. Joint spatio-temporal modelling of adverse pregnancy outcomes sharing common risk factors at sub-county level in Kenya, 2016–2019. BMC Public Health. 2021;21:1–13.
- 21. Afeworki R, Smits J, Tolboom J, van der Ven A. Positive effect of large birth intervals on early childhood hemoglobin levels in Africa is limited to girls: cross-sectional DHS study. PLoS ONE. 2015;10:e0131897.
- 22. Budu E, Ahinkorah BO, Ameyaw EK, Seidu A-A, Zegeye B et al. (2021) Does Birth Interval Matter in Under-Five Mortality? Evidence from Demographic and Health Surveys from Eight Countries in West Africa. BioMed Research International 2021: 5516257.
- Molitoris J, Barclay K, Kolk M. When and where birth spacing matters for child survival: an international comparison using the DHS. Demography. 2019;56:1349–70.
- Usama M, Qadir J, Raza A, Arif H, Yau K-LA, et al. Unsupervised machine learning for networking: techniques, applications and research challenges. IEEE Access. 2019;7:65579–615.
- Hajian-Tilaki K, Asnafi N, Aliakbarnia-Omrani F. The patterns and determinants of birth interval in multiparous women in Babol, Northern Iran. Southeast Asian J Trop Med Public Health. 2009;40:852.
- 26. Shan H, Gubin El. Data cleaning for data analysis; 2019. pp. 387–8.
- 27. Suthaharan S. Machine learning models and algorithms for big data classification. Integr Ser Inf Syst. 2016;36:1–12.
- Yehuala TZ, Derseh NM, Tewelgne MF, Wubante SM. (2024) Exploring machine learning algorithms to predict diarrhea disease and identify its determinants among Under-Five years children in East Africa. J Epidemiol Global Health: 1–11.
- 29. Deeba F, Patil SR. Utilization of machine learning algorithms for prediction of diseases; 2021. IEEE. pp. 1–7.
- Poola I. (2017) The Best of the Machine Learning Algorithms Used in Artificial Intelligence. International Journal of Advanced Research in Computer and Communication Engineering ISO 3297: 2007 Certified 6.

- 32. Haight SC, Hogue CJ, Raskind-Hood CL, Ahrens KA. Short interpregnancy intervals and adverse pregnancy outcomes by maternal age in the united States. Ann Epidemiol. 2019;31:38–44.
- 33. Johnson K, Posner SF, Biermann J, Cordero JF, Atrash HK et al. (2006) Recommendations to improve preconception health and Health Care—United States: report of the CDC/ATSDR preconception care work group and the select panel on preconception care. Morbidity and Mortality Weekly Report: Recommendations and Reports 55: 1-CE-4.
- Merklinger-Gruchala A, Jasienska G, Kapiszewska M. Short interpregnancy interval and low birth weight: A role of parity. Am J Hum Biology. 2015;27:660–6.
- 35. Ningrum RA, Fahmiyah I, Levi A, Syahputra MA. Short birth intervals classification for Indonesia's women. Bull Electr Eng Inf. 2022;11:1535–42.
- Seidman DS, Armon Y, Roll D, Stevenson DK, Gale R. Grand multiparity: an obstetric or neonatal risk factor? Am J Obstet Gynecol. 1988;158:1034–9.
- Fayehun O, Omololu O, Isiugo-Abanihe UC. Sex of preceding child and birth spacing among Nigerian ethnic groups. Afr J Reprod Health. 2011;15:79–89.
- Hailu D, Gulte T. (2016) Determinants of short interbirth interval among reproductive age mothers in Arba minch district, Ethiopia. International journal of reproductive medicine 2016.
- 39. Ahmed S, Creanga AA, Gillespie DG, Tsui AO. Economic status, education and empowerment: implications for maternal health service utilization in developing countries. PLoS ONE. 2010;5:e11190.

- 40. Ganatra B, Faundes A. Role of birth spacing, family planning services, safe abortion services and post-abortion care in reducing maternal mortality. Best Pract Res Clin Obstet Gynecol. 2016;36:145–55.
- Pasha O, Goudar SS, Patel A, Garces A, Esamai F, et al. Postpartum contraceptive use and unmet need for family planning in five low-income countries. Reproductive Health. 2015;12:1–7.
- Brown W, Ahmed S, Roche N, Sonneveldt E, Darmstadt GL. Impact of family planning programs in reducing high-risk births due to younger and older maternal age, short birth intervals, and high parity. Elsevier; 2015. pp. 338–44.
- 43. Alemayehu T, Haider J, Habte D. (2010) Determinants of adolescent fertility in Ethiopia. Ethiop J Health Dev 24.
- Hailu D, Gulte T. Determinants of short interbirth interval among reproductive age mothers in Arba minch district, Ethiopia. Int J Reproductive Med. 2016;2016:6072437.
- 45. Smits Z, Jongbloet B. The association of birth interval, maternal age and season of birth with the fertility of daughters: a retrospective cohort study based on family reconstitutions from nineteenth and early twentieth century Quebec. Paediatr Perinat Epidemiol. 1999;13:408–20.

Publisher's note

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