


Explainable machine learning for modeling predictors of unintended pregnancy among married/in-union women in sub-Saharan Africa, a multi-country analysis of MICS 6 survey

DIGITAL HEALTH
Volume 10: 1–10
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241292978
journals.sagepub.com/home/dhj



Shimels Derso Kebede¹ , Kaleab M Abera², Eyob Tilahun Abeje³, Ermias Bekele Enyew¹, Chala Daba⁴, Lakew Asmare³, Fekade Demeke Bayou³, Mastewal Arefaynie⁵, Anissa Mohammed³, Abiyu Abadi Tareke⁶, Awoke Keleb⁴, Natnael Kebede⁷, Yawkal Tsega² and Abel Endawkie³

Abstract

Introduction: Unintended pregnancy is defined as a pregnancy that is either mistimed (wanted at a later time) or unwanted (not wanted at all). It has been a concerning issue for reproductive health and public health, with significant negative effects on the mother, child, and the public at large. It is a worldwide public health issue that can have a major impact on the health of pregnant women and newborns.

Methods: The study was conducted using secondary data from IPUMS Multiple Indicator Cluster Surveys round 6. The analysis was based on a data merged from six sub-Saharan Africa countries such as Gambia, Ghana, Lesotho, Malawi, Nigeria, and Sierra Leone. A total weighted sample of 28,027 married/in-union reproductive-age women was included in the study. Seven machine learning algorithms were trained and their performance compared in predicting unintended pregnancy. Finally, Shapley Additive exPlanations model explanation technique was used to identify the predictors of unintended pregnancy.

Results: XGBoost was the top-performing model, achieved the highest area under receiver operating characteristic curve (0.62) and accuracy (65.92%), surpassing all other models. SHAP global feature importance identified top predictors of unintended pregnancy, with women from Malawi, Ghana, and Lesotho, women having primary education and secondary education, with parity of more than three, have higher likelihood of unintended pregnancy. In the other hand, women from Nigeria and Sierra Leone, whose husband/partner has more wives or partners (polygamy relation), and women who owns mobile phone had lower risk of unintended pregnancy.

Conclusion: These findings highlight the importance of considering contextual factors, such as country-specific sociocultural norms and individual characteristics, in understanding and addressing unintended pregnancies. By strategically addressing the identified predictors, policymakers, and healthcare providers can develop impactful programs that address the root causes of unintended pregnancies, ultimately contributing to improved reproductive health outcomes worldwide.

¹Department of Health Informatics, School of Public Health, College of Medicine and Health Science, Wollo University, Dessie, Ethiopia

²Department of Health System and Management, School of Public Health, College of Medicine and Health Sciences, Wollo University, Dessie, Ethiopia

³Department of Epidemiology and Biostatistics, School of Public Health, College of Medicine and Health Sciences, Wollo University, Dessie, Ethiopia

⁴Department of Environmental Health, College of Medicine and Health Sciences, Wollo University, Dessie, Ethiopia

⁵Department of Reproductive and Family Health, School of Public Health, College of Medicine and Health Sciences, Wollo University, Dessie, Ethiopia

⁶Amref Health in Africa, COVID-219 Vaccine/EPI Technical Assistant at West Gondar Zonal Health Department, Gondar, Ethiopia

⁷Department of Health Promotion, School of Public Health, College of Medicine and Health Sciences, Wollo University, Dessie, Ethiopia

Corresponding author:

Shimels Derso Kebede, Department of Health Informatics, School of Public Health, College of Medicine and Health Science, Wollo University, Dessie, Ethiopia.

Email: shime4d@gmail.com



Keywords

Unintended pregnancy, MICS, machine learning, explainable AI, sub-Saharan Africa, Shapley value, SHAP

Submission date: 29 February 2024; Acceptance date: 28 August 2024

Introduction

Unintended pregnancy is defined as a pregnancy that is either mistimed (wanted at a later time) or unwanted (not wanted at all).¹ It has been a concerning issue for reproductive health and public health, with significant negative effects on the mother, child, and the public at large. It is a worldwide public health issue that can have a major impact on the health of pregnant women and newborns.^{2,3}

Worldwide, an estimated 44% of pregnancies were unintended in 2010–2014. Between 2010 and 2014, 59% of unwanted births in developed areas and 55% of unwanted births in developing areas resulted in abortions. And 38% of unplanned pregnancies in Africa result in an abortion. Recently, the United Nations sexual and reproductive health agency reported that, nearly half of all pregnancies are unintended globally.⁴ The unintended pregnancy rate remains significantly higher in developing countries than in developed countries.⁵ Africa has an unintended pregnancy rate of 89 per 1000 women aged 15 to 44. In sub-Saharan Africa (SSA), the prevalence of unintended pregnancies was 29.0% ranged from 10.8% in Nigeria to 54.5% in Namibia.⁶ Unwanted pregnancies result in 47,000 maternal deaths and 25 million unsafe abortions, every year.^{7,8}

Unintended pregnancies can result in maternal mental illness and are associated with increased maternal and neonatal morbidity and mortality, often associated with late initiation and inadequate use of antenatal care services,⁹ tobacco or alcohol use during pregnancy, low birth weight infants, and decreased breastfeeding rates.¹⁰ Unintended pregnancies can also result in a number of health dangers, including illnesses, malnourishment, abuse, intimate partner violence, depression, suicidal ideation, anxiety, stress, lower relationship satisfaction, and social support.^{11,12} Moreover, unwanted pregnancies result in cycles of high fertility, poverty, and reduced opportunities for education and employment problems. Children born from unintended pregnancies may experience cognitive and physical disabilities, lower educational attainment, and have lower self-esteem as young adults¹³ and unplanned and mistimed children exhibit more behavioral disorders.¹⁴

Previous studies have identified numerous predictors such as age, level of education, marital status, parity and region of residence, wealth status, had ever given birth, their sexual debut, as significant predictors of unintended pregnancies.^{15–18} Though unintended pregnancy predicted by those factors, it may also associated by other social, cultural and economic factors across variety of nations. Hence machine learning is a better analysis to reveal previously unseen relationships. Moreover, machine learning helps to develop predictive models and to rank important predictors based on their effect on the outcome variable. This study aimed to identify novel insights into the correlates of unintended pregnancy by applying machine learning methods to model multinational survey data and applying Shapley Additive exPlanations (SHAP) model explanation technique to interpret model predictions.

Method

Data source

The study was conducted using secondary data from IPUMS Multiple Indicator Cluster Surveys (MICS) round 6¹⁹ that was downloaded from the website (<https://mics.unicef.org/surveys>). IPUMS MICS is the integrated version of UNICEF MICS, which is the largest and most comprehensive source of data on women and children's health across the world, including countries in Africa, Eastern Europe, Asia, and Latin America.²⁰ Most of MICS surveys are nationally representative and an integral part of plans and policies of many governments around the world, and a major data source for more than 30 Sustainable Development Goals (SDGs) indicators. The information obtained through MICS surveys constitutes topics in maternal and child health, education and child mortality to child protection, HIV/AIDS and water and sanitation.

The dataset used for this study was MICS 6 cross-sectional survey data from six SSA countries such as Sierra Leone (2017), Ghana (2017/18), Gambia (2018), Lesotho (2018), Malawi (2019/20), and Nigeria (2021). The survey instrument used is the standardized MICS 6 questionnaire, which is a widely recognized and validated tool for collecting data on various indicators related to child health and development. The MICS 6 questionnaire

is publicly available and can be accessed at https://mics.ipums.org/mics/resources/enum_materials_pdf/survey_form_mics6_wm.pdf.

Inclusion and exclusion criteria

Women between the ages of 15 and 49 years who are married or living with a partner were considered to meet the inclusion criteria. Women from countries that doesn't collect on the dependent variable in their MICS6 survey were excluded in this study.

Study variables

Outcome variable. Unintended pregnancy was the dependent variable which was dichotomized into two categories such as “intended” and “unintended” following the definition of CDC,²¹ that defined unintended pregnancy is pregnancy that is either wanted then or later than occurred (mistimed) or not needed (unwanted). So we coded pregnancy wanted then = 0 “intended”; later and not at all = 1 “unintended.”⁶

Predictors. We have included women age, health insurance coverage, women's education level, wealth index, place of residence, country, media exposure, ever used internet, mobile ownership, parity, wealth index, polygamy relation, substance use, overall life happiness, age at first sex, history of child death, and history of family planning method use as potential predictors of unintended pregnancy. These predictor variables were selected from previous studies investigated unintended pregnancy.^{6,15–18,22–24}

Data processing and analysis

A total weighted sample of 28,027 married/in-union reproductive-age women was included for analysis from these 6 countries. Then the data undergone various preprocessing tasks such as data cleaning, feature engineering, and data splitting.

During data cleaning we have imputed missing values in the predictor variables through “CALIBERrfimpute” package in R. CALIBERrfimpute imputes missing values using random forest under full conditional specifications (multivariate imputation by chained equations).²⁵ The choice of this method was based on the assumption that the missing data mechanism in our dataset was missing completely at random (MCAR). Hence, the imputation process would not introduce any systematic bias, and the results can be considered unbiased and robust. Unbalanced categories of the outcome variable was balanced using Synthetic Minority Oversampling Technique (SMOTETomek) to avoid machine learning models biased toward the majority class. Another data cleaning task carried out was One-Hot-Encoding to

encode categorical variables to dummy variables with each category as a separate variable. Finally, the whole data was randomly divided into two sets; such that 80% of the data was training set for model training and the remaining 20% testing for evaluating model performance.

The data used in this study were concurrent, meaning the observations were collected independently of each other over a period without any inherent order. Given the concurrent nature of our data, we have selected popular classification algorithms to predict unintended pregnancy. A total of seven machine learning algorithms such as LightGBM, Catboost, XGBoost, RandomForest, AdaBoost, Logistic regression, and Pytorch_Tabular (a deep learning model for tabular data). The selected models were trained on balanced and unbalanced data and their performance was compared. During training each model was trained on default hyperparameter and tuned hyperparameters, as hyperparameter tuning was done for all models through Optuna framework. Then, the performance of the models was compared through classification matrices such as accuracy and area under receiver operating characteristic curve (AUC) score. Finally, the best model was selected and used for identifying important predictors of unintended pregnancy, through SHAPs global and local explanations.

Results

Characteristics of respondents

The descriptive statistics revealed details about women across different characteristics (Table 1). In terms of age distribution, the largest group, comprising 44.64%, fell into the 26 to 35 age range, followed by 37.0% in the 15 to 25 category, and 18.36% in the 36 to 49 group. Regarding health insurance, a significant majority (91.45%) lacked coverage, while only a small percentage (8.55%) was insured. Educational levels varied, with 37.57% having no formal education, 25.63% completing primary schooling, and 7.13% attaining higher education. Media exposure was reported by 60.21% of women, contrasting with 39.79% having no exposure. Internet access was limited (13.21%), while mobile ownership was evenly distributed (48.79% owned a mobile device). The majority (58.15%) of women had 1 to 3 children, and 41.85% had more than 3. In terms of wealth, 46.04% were classified as poor, 19.01% as middle class, and 34.95% as rich. Substance use was low (8.11%), with 91.89% reporting none. Residentially, 63.97% lived in rural areas, and 36.03% in urban areas. Regarding the age of first sexual experience, 59.50% experienced it before 18, while 40.50% did so after turning 18. Family planning practices were reported by 20.24%, while 79.76% did not use any methods. The data encompassed multiple countries, with Nigeria (33.06%), Sierra Leone (23.80%), and Malawi

Table 1. Characteristics of respondents.

Variable	Categories	Weighted frequency	Percentage
Age of woman	15 to 25	10,369	37.00
	26 to 35	12,512	44.64
	36 to 49	5146	18.36
Health insured	No	24,864	91.45
	Yes	2326	8.55
Education level	No formal	10,529	37.57
	Primary	7182	25.63
	Secondary	8316	29.67
	Higher	1998	7.13
Media exposure	No	11,152	39.79
	Yes	16,874	60.21
Internet access	No	23,820	86.79
	Yes	3626	13.21
Mobile ownership	No	14,332	51.21
	Yes	13,656	48.79
Parity	One to three	16,297	58.15
	More than three	11,730	41.85
Wealth index	Poor	12,903	46.04
	Middle	5328	19.01
	Rich	9796	34.95
Polygamy relation	No	21,197	75.88
	Yes	6740	24.12
Substance use	No	25,753	91.89
	Yes	2274	8.11
Overall life happiness	Happy	21,072	75.23
	Neutral	4603	16.43
	Unhappy	2337	8.34

(continued)

Table 1. Continued.

Variable	Categories	Weighted frequency	Percentage
Place of residence	Urban	10,097	36.03
	Rural	17,929	63.97
Age at first sex	Before 18	13,733	59.50
	After 18	9349	40.50
History of child death	No	21,607	77.10
	Yes	6419	22.90
Ever used FP method	No	15,126	79.76
	Yes	3839	20.24
Country	Gambia	3244	11.58
	Ghana	2928	10.45
	Lesotho	836	2.98
	Malawi	5083	18.13
	Nigeria	9266	33.06
	Sierra Leone	6670	23.80

(18.13%) being the most represented, while Gambia, Ghana, and Lesotho had smaller proportions in the dataset.

Model performance comparison

We conducted a comparison of various machine learning models both before and after balancing the training data, employing default and tuned hyperparameters. In the case of unbalanced training data, prior to hyperparameter tuning, the models exhibited varying levels of performance, with AUC values ranging from 0.54 to 0.60 and accuracy ranging from 70.11% to 72.66% and after hyperparameter tuning, some models showed slight improvements in performance (Table 2). LightGBM, CatBoost, and XGBoost all had similar AUC values of around 0.57, with CatBoost achieving the highest accuracy of 72.8%. Random Forest and Logistic Regression models also had similar AUC values, but their accuracy remained relatively unchanged. When the dataset was balanced, a different pattern emerged. The AUC values increased for most models, ranging from 0.58 to 0.62, indicating improved model performance. LightGBM and CatBoost also showed improved AUC values, but their accuracy

Table 2. Model comparison.

Model	Before balancing			
	Before hyper parameter tuning		After hyper parameter tuning	
	AUC	Accuracy	AUC	Accuracy
LightGBM	0.56	71.94	0.56	72.62
Catboost	0.56	72.66	0.57	72.8
XGBoost	0.56	71.74	0.57	71.82
Random forest	0.57	70.11	0.53	72
AdaBoost	0.54	72.12	0.54	71.94
Logistic regression	0.54	71.30	0.54	71.35
PyTorch tabular	0.60	71.00	0.6	71.1
After balancing				
LightGBM	0.61	63.42	0.61	63.00
CatBoost	0.61	63.58	0.61	63.29
XGboost	0.60	63.06	0.62	65.92
Random forest	0.58	62.28	0.61	64.71
Adaboost	0.61	62.15	0.61	60.86
Logistic regression	0.61	62.33	0.61	62.26
PyTorch tabular	0.61	60.36	0.61	60.4

decreased slightly. Notably, XGBoost achieved the highest AUC (0.62) and accuracy (65.92%) after balancing, surpassing all other models.

Predictors of unintended pregnancy

SHAP global feature importance was used for selecting top predictors of unintended pregnancy based on feature's contribution toward the dependent variable. Accordingly, Malawi (country_4), Nigeria (country_5), Ghana (country_2), women secondary education (women_education_2), parity of more than 3 (parity_2), Husband/partner has more wives or partners (polygamy_relation_1), Sierra Leone (country_6), primary women education (women_education_1), Lesotho (country_3), and owning a mobile phone (own_mobilePhone_1) were top predictors of unintended pregnancy (Figure 1).

The relationship between the predictors and the outcome variable was further explained by SHAP Beeswarms plot to provide a clear explanation of how these predictor variables impact the model prediction globally. In Figure 2, the x -axis represents Shapley values of each variable on the model output computed from XGBoost model. Every instance (row) of the dataset appears as its own point for each variable. The points are distributed horizontally along the x -axis according to their SHAP value. Examining how the SHAP values are distributed reveals how a variable influence the model predictions. High/positive SHAP values are associated with an increase in the predicted probability, whereas low/negative SHAP values correspond to a reduction in the predicted probability. Hence, dots that are right to the vertical line (0 SHAP value) increase the likelihood of unintended pregnancy while the left side decreases unintended pregnancy likelihood. The color bar on the right side corresponds to the raw values of the variables for each instance on the graph. If the value of a variable for a particular instance is relatively high, it appears as a red point. Relatively low variable values appear as blue points. In our case, the red indicates higher value (1) and blue indicates lower value (0) since all variables are categorical with binary value (0 or 1). Accordingly, women from Malawi (country_4), Ghana (country_2), and Lesotho (country_3), women having primary education (women_education_1), and secondary education (women_education_2), with parity of more than three (parity_2), have higher likelihood of unintended pregnancy (Figure 2). On the other hand, women from Nigeria (country_5) and from Sierra Leone (country_6), whose husband/partner has more wives or partners (polygamy_relation_1), and women who owns mobile phone had lower risk of unintended pregnancy.

The distribution of points is also informative. For country_4 (Malawi), we see a dense cluster of low value (blue points) with small negative SHAP values. Instances of higher values (red points) extend further toward the left, suggesting higher values of this variable (i.e. country_4 = 1) has a stronger positive impact on unintended pregnancy than a negative impact of lower values of this variable (i.e. country_4 = 0).

Discussion

The study aimed to examine the predictors of unintended pregnancy using machine learning models. The analysis of machine learning models for predicting unintended pregnancy revealed interesting findings. The models' performance varied before and after balancing the training data. After balancing, most models showed improved performance, with increased AUC values indicating better predictive ability. XGBoost emerged as the top-performing model, achieving the highest AUC and accuracy after balancing the dataset. This suggests that XGBoost has the potential to

accurately predict unintended pregnancies for this specific dataset.

In our study, we acknowledge that the preprocessing steps of missing value imputation and data balancing may influence the results. In utilizing CALIBERrfimpute for missing value imputation and SMOTETomek for class balancing, it's important to consider their potential impacts on our study's outcomes. The imputation of missing values with CALIBERrfimpute may influence the dataset's distribution and statistical properties, potentially affecting downstream analyses and model performance. Similarly, SMOTETomek alters the class distribution by synthesizing minority class samples and removing borderline majority class instances, which could introduce biases or affect the model's predictive capabilities.

The identification of top predictors of unintended pregnancy using SHAP global feature importance provides valuable insights regarding top predictors of unintended pregnancy in Africa. The country of residence, education level, parity, polygamy relation, and mobile phone ownership were identified as significant predictors. The SHAP model explanation further elucidates the relationship between the predictors and unintended pregnancy. It demonstrates that women from certain countries, such as Malawi, Ghana, and Lesotho have a higher likelihood of unintended pregnancy.

Similarly, lower levels of education were related with higher likelihood of unintended pregnancy. This finding was in line with studies in Ethiopia^{26,27} and Ghana,²⁸ reported a higher probability of unintended pregnancy among women with lower level of education. This might be due to uneducated women is less likely to know about contraceptive methods and less aware of contraceptive choices, and having lower awareness of contraception significantly associates with higher level of unintended pregnancy.²⁹ Another explanation for this may be women with lower level of education are low empowered to take control of their sexual and reproductive health matters.

In this study, higher parity was also related with a higher probability of unintended pregnancy. This finding is consistent with findings in studies in Ethiopia^{3,30,31} and Ghana.²⁸ The possible explanation might be these women have enough or adequate number of children with a decreasing intention for the next pregnancy or childbirth. Another possible reason may be the fact that fertility preference among multipara women is lower than nulliparous women,³¹ which may cause pregnancy among multiparous women to be unintended.

On the other hand, women from Nigeria and Sierra Leone, those in polygamous relationships were associated with a lower probability of unintended pregnancy. Women whose husband/partner has more wives or partners had a lower probability of unintended pregnancy. This finding confirms the findings of a study in Ethiopia reported a lower level of experience unintended pregnancy in

women whose husbands were in a polygamy relationship.²³ Similarly, women who owns a mobile phone had a lower likelihood of unintended pregnancy. This finding was supported by a study conducted in Democratic Republic of the Congo.³² This might be due to owning a mobile phone exposes women to access vital information regarding family planning, maternal and child health, and their overall healthcare. Moreover, owning mobile phones provides a means of communication with family and friends, improves women empowerment.

In addition to these factors, it is important to consider the prevalence and type of contraceptive measures in a population can significantly influence reproductive health outcomes such as unintended pregnancy. In our study, the analyzed countries exhibit varying levels of modern contraceptive use ranging from 6.5% in Gambia to 48.5% in Lesotho as of a study report in 2022.³³ The differences in contraceptive use across the analyzed countries may partially explain the observed variations in reproductive health outcomes. Higher contraceptive prevalence is generally associated with better maternal health indicators and lower rates of unintended pregnancies. In countries with lower contraceptive use, the higher incidence of unintended pregnancies can lead to increased health risks for mothers and children, affecting overall health outcomes.

Our study differs from previous research in several key aspects, including the use of explainable machine learning techniques to model predictors of unintended pregnancy. Unlike traditional regression approaches, explainable machine learning provides insights into the complex interactions among variables, enhancing the interpretability of our predictive models. Furthermore, our multi-country analysis using MICS 6 survey data allows for cross-country comparisons and insights into regional variations, which have not been extensively explored in prior literature. By employing advanced machine learning techniques and integrating a large-scale survey dataset, our study advances methodological approaches to studying unintended pregnancies in sub-Saharan Africa. We highlight the strengths and limitations of our approach compared to existing methodologies, underscoring the innovative aspects of our research design.

Strength and limitations of the study

The present study has some limitations. First, the study included data from countries that provided data in English language. While the MICS 6 survey provides a valuable dataset for our analysis, it is important to note that the data are self-reported and subject to recall bias. Additionally, certain variables of interest may have limited availability or granularity across different countries, which could impact the depth of our analysis. Second, as oversampling of minority class was done to balance the training data the generalizability of the models may be affected.

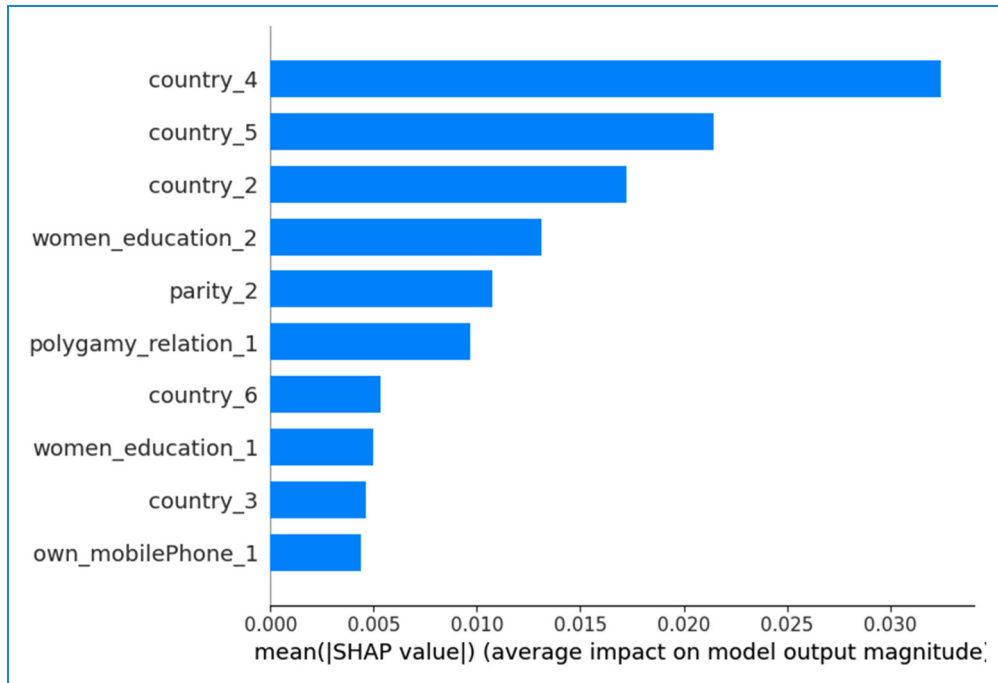


Figure 1. Top predictors of unintended pregnancy based on SHAP global feature importance computed by XGBoost model (country_4 = Malawi, country_5 = Nigeria, country_2 = Ghana, women_education_2 = women secondary education, parity_2 = parity of more than 3, polygamy_relation_1 = Husband/partner has more wives or partners, country_6 = Sierra Leone, women_education_1 = primary women education, country_3 = Lesotho, own_mobilePhone_1 = owning a mobile phone). SHAP: Shapley Additive exPlanations.

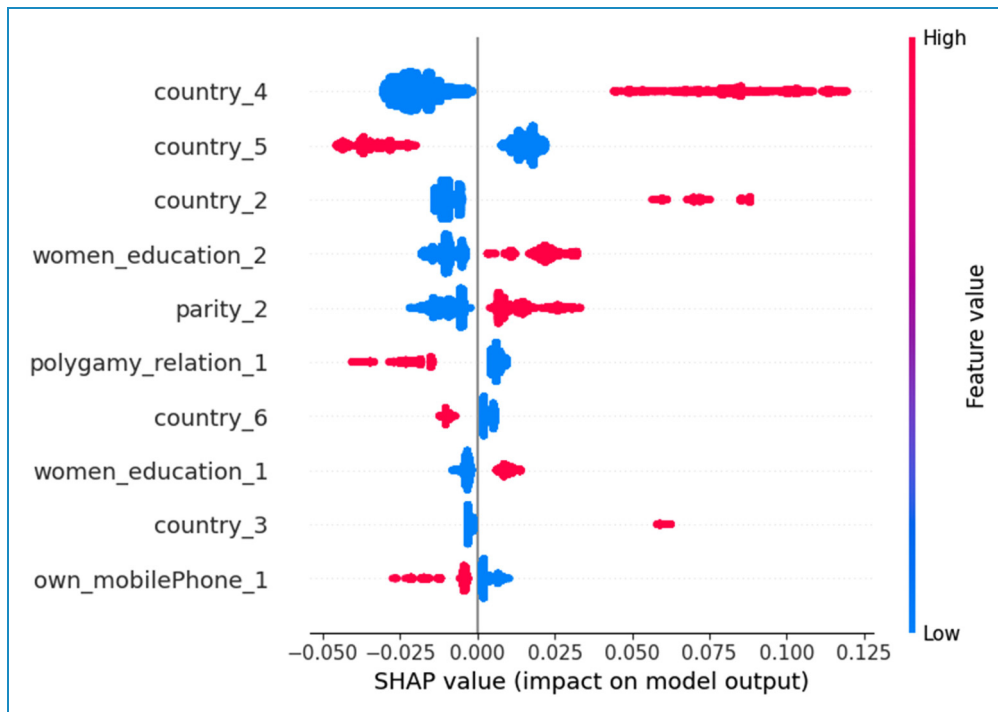


Figure 2. SHAP Beeswarm plot based on XGBoost model. (country_4 = Malawi, country_5 = Nigeria, country_2 = Ghana, women_education_2 = women secondary education, parity_2 = parity of more than 3, polygamy_relation_1 = Husband/partner has more wives or partners, country_6 = Sierra Leone, women_education_1 = primary women education, country_3 = Lesotho, own_mobilePhone_1 = owning a mobile phone). SHAP: Shapley Additive exPlanations.

Despite these limitations, the present study makes an important scientific contribution filling the information gap regarding predictors of unintended pregnancy among married/in-union women in sub-Saharan Africa. Furthermore, this study used explainable artificial intelligence (XAI) techniques to explore the effect of identified predictors on the outcome variable to address the noninterpretability limitation of black-box machine learning models.

Future studies may conduct longitudinal studies to provide insights into the dynamic nature of predictors of unintended pregnancy over time, allowing for more robust causal inference. Additionally, supplementing quantitative analyses with qualitative research methods could offer deeper insights into the sociocultural factors influencing unintended pregnancies and contraceptive use.

Conclusions

This study was aimed to predict and identify predictors of unintended pregnancy. The in-depth has uncovered critical insights into the multifaceted nature of this issue. The meticulous comparison of machine learning models accentuated not only their efficacy but also underscored the significance of data balancing in enhancing predictive accuracy. Notably, XGBoost emerged as the top-performing model, showcasing its potential as a robust tool for predicting unintended pregnancies. This revelation provides a practical avenue for implementing predictive models in public health efforts to identify and support populations at higher risk.

The utilization of SHAP global feature importance added a layer of granularity to the analysis, spotlighting specific predictors with substantial influence on unintended pregnancies. Notably, the country of residence emerged as a pivotal factor, emphasizing the need for geographically tailored interventions. Moreover, education level, parity, polygamy relations, and mobile phone ownership surfaced as influential predictors. These variables provide actionable insights for targeted interventions. Recognizing the role of education and family planning in reducing unintended pregnancies suggests the importance of investing in educational programs and accessible healthcare services. Addressing cultural aspects such as polygamy relations underscores the need for culturally sensitive approaches, while leveraging mobile phone ownership implies the potential for innovative outreach and communication strategies. Additionally, the identified predictors highlight the importance of region-specific strategies, emphasizing the need for nuanced approaches to address diverse sociocultural contexts.

In conclusion, this comprehensive analysis not only advances our academic understanding of unintended pregnancies but also provides practical recommendations for targeted interventions. By strategically addressing the identified predictors, policymakers, and healthcare providers

can develop impactful programs that address the root causes of unintended pregnancies, ultimately contributing to improved reproductive health outcomes worldwide.

List of abbreviations

AUC	area under receiver operating characteristic curve
MICS	Multiple Indicator Cluster Surveys
SDG	Sustainable Development Goals
SHAP	Shapley Additive exPlanation
SMOTETomek	Synthetic Minority Oversampling Technique
SSA	sub-Saharan Africa
XAI	explainable artificial intelligence.

Acknowledgments: The authors are grateful to the UNICEF's MICS survey project for making the data publicly available. We are also grateful for the open source python community.

Availability of data and materials: The data is available on the public domain through the UNICEF website (<https://mics.unicef.org>)

Contributorship: SDK was responsible for a significant contribution to the work reported in the conceptualization, data preparation, analysis & interpretation, and original draft preparation. KM, ETA, EBE, CD, LA, FDB, MA, AM, AAT, AK, NK, YT, and AE have made significant contributions to data preparation, drafting, revising, or critically reviewing the article; and agree to be accountable for all aspects of the work. The final manuscript of the work was read, edited, and approved by all authors.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval and consent to participate: There was no need for ethical clearance, as the study used publicly available data which are anonymized and researchers did not interact with study participants. The MICS6 survey was approved by the relevant institutional and/or licensing committee in each country where the survey was conducted and informed consent was obtained from the respondents or their guardians before their participation in the survey. All data used in this study were de-identified to ensure the privacy and confidentiality of the participants. Personal identifiers were removed, and unique codes were assigned to each data entry to prevent any potential re-identification. Furthermore, permission to use the data has been granted by UNICEF, after registration to access the dataset.

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

Guarantor: SDK

ORCID ID: Shimels Derso Kebede  <https://orcid.org/0000-0001-5416-616X>

References

- Reich JA. Reconceiving unintended pregnancy. In: *Introducing the new sexuality studies: original essays*. 4th ed. London: Routledge, 2022, pp.143–152.
- Getu Melese K, Gebrie MH, Berta Badi M, et al. Unintended pregnancy in Ethiopia: community based cross-sectional study. *Obstet Gynecol Int* 2016; 2016: 1–5.
- Ayele A, Abdurashid N, Hailu M, et al. Unintended pregnancy and associated factors among pregnant women attending antenatal care unit in public health facilities of Dire Dawa City, Eastern Ethiopia, 2021. *Obstet Gynecol Int* 2023; 2023: 8100462.
- Nearly half of all pregnancies are unintended—a global crisis, says new UNFPA report | United Nations Population Fund, <https://www.unfpa.org/press/nearly-half-all-pregnancies-are-unintended-global-crisis-says-new-unfpa-report>.
- Bearak J, Popinchalk A, Alkema L, et al. Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a Bayesian hierarchical model. *Lancet Glob Health* 2018; 6: e380–e389.
- Ameyaw EK, Budu E, Sambah F, et al. Prevalence and determinants of unintended pregnancy in sub-Saharan Africa: a multi-country analysis of demographic and health surveys. *PLoS ONE* 2019; 14: e0220970.
- Blencowe H, Cousens S, Jassir FB, et al. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *Lancet Glob Health* 2016; 4: e98–e108.
- Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet* 2016; 388: 3027–3035.
- Dibaba Y, Fantahun M and Hindin MJ. The effects of pregnancy intention on the use of antenatal care services: systematic review and meta-analysis. *Reprod Health* 2013; 10: 1–9.
- Preventing unintended pregnancy: the contraceptive CHOICE project in review. *J Womens Health (Larchmt)* 2015; 24: 349–353.
- Faisal-Cury A, Menezes PR, Quayle J, et al. Unplanned pregnancy and risk of maternal depression: secondary data analysis from a prospective pregnancy cohort. *Psychol Health Med* 2017; 22: 65–74.
- Yanikkerem E, Ay S and Piro N. Planned and unplanned pregnancy: effects on health practice and depression during pregnancy. *J Obstet Gynaecol Res* 2013; 39: 180–187.
- World Health Organization. High rates of unintended pregnancies linked to gaps in family planning services: New WHO study. [updated 2019/10/25/], <https://www.who.int/news/item/25-10-2019-high-rates-of-unintended-pregnancies-linked-to-gaps-in-family-planning-services-new-who-study> (2019).
- Carson C, Redshaw M, Sacker A, et al. Effects of pregnancy planning, fertility, and assisted reproductive treatment on child behavioral problems at 5 and 7 years: evidence from the Millennium Cohort Study. *Fertil Steril* 2013; 99: 456–463.
- Ayalew HG, Liyew AM, Tessema ZT, et al. Prevalence and factors associated with unintended pregnancy among adolescent girls and young women in sub-Saharan Africa, a multi-level analysis. *BMC Womens Health* 2022; 22: 464.
- Ameyaw EK. Prevalence and correlates of unintended pregnancy in Ghana: analysis of 2014 Ghana Demographic and Health Survey. *Matern Health Neonatol Perinatol* 2018; 4: 1–6.
- Grindlay K, Dako-Gyeke P, Ngo TD, et al. Contraceptive use and unintended pregnancy among young women and men in Accra, Ghana. *PLoS ONE* 2018; 13: e0201663.
- Nyarko SH. Unintended pregnancy among pregnant women in Ghana: prevalence and predictors. *J Pregnancy* 2019; 2019:1–8.
- Bolgrien A, Boyle EH, Sobek M, et al. *IPUMS-MICS data harmonization code: version 1.1*. Minneapolis, MN: IPUMS, 2024.
- Khan S and Hancioglu A. Multiple indicator cluster surveys: delivering robust data on children and women across the globe. *Stud Fam Plann* 2019; 50: 279–286.
- CDC. Unintended Pregnancy | CDC [updated 2023/06/15/T09:46:45Z], <https://www.cdc.gov/reproductivehealth/contraception/unintendedpregnancy/index.htm> (2023).
- Sarder A, Islam SMS, Maniruzzaman, et al. Prevalence of unintended pregnancy and its associated factors: evidence from six South Asian countries. *PLoS ONE* 2021; 16: e0245923.
- Zelege LB, Alemu AA, Kassahun EA, et al. Individual and community level factors associated with unintended pregnancy among pregnant women in Ethiopia. *Sci Rep* 2021; 11: 12699.
- Teshale AB and Tesema GA. Magnitude and associated factors of unintended pregnancy in Ethiopia: a multilevel analysis using 2016 EDHS data. *BMC Pregnancy Childbirth* 2020; 20: 329.
- Shah AD, Bartlett JW, Carpenter J, et al. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *Am J Epidemiol* 2014; 179: 764–774.
- Merga J, Wirtu D, Bekuma TT, et al. Unintended pregnancy and the factors among currently pregnant married youths in Western Oromia, Ethiopia: a mixed method. *PLoS ONE* 2021; 16: e0259262.
- Zeru MA, Fenta HM and Mitku AA. Spatial patterns and predictors of unintended pregnancy among reproductive age women in Ethiopia. *PLoS ONE* 2023; 18: e0282225.
- Ahinkorah BO, Seidu AA, Appiah F, et al. What has reproductive health decision-making capacity got to do with unintended pregnancy? Evidence from the 2014 Ghana Demographic and Health Survey. *PLoS ONE* 2019; 14: e0223389.
- Mohammed F, Musa A and Amano A. Prevalence and determinants of unintended pregnancy among pregnant woman attending ANC at Gelemso General Hospital, Oromiya Region, East Ethiopia: a facility based cross-sectional study. *BMC Womens Health* 2016; 16: 56.
- Bekele H, Dheressa M, Mengistie B, et al. Unintended pregnancy and associated factors among pregnant women attending antenatal care at Bako Tibe District Public Health Facility, Oromia Region, Ethiopia. *J Pregnancy* 2020; 2020: 3179193.

-
31. Bekele YA and Fekadu GA. Factors associated with unintended pregnancy in Ethiopia; further analysis of the 2016 Ethiopian demographic health survey data. *BMC Pregnancy Childbirth* 2021; 21: 486.
 32. Dhakal S, Song JS, Shin DE, et al. Unintended pregnancy and its correlates among currently pregnant women in the Kwango District, Democratic Republic of the Congo. *Reprod Health* 2016; 13: 74.
 33. Boadu I. Coverage and determinants of modern contraceptive use in sub-Saharan Africa: further analysis of demographic and health surveys. *Reprod Health* 2022; 19: 18.
-