



Socio-Demographic Predictors of Maternal Health-Related Outcomes: A Systematic Review of Machine Learning Studies in Sub-Saharan Africa

***Sandra Kemboi, Emma Anyika & Peter Muturi**

Department of Computer Science and Information Technology, School of Computing and Mathematics,
The Co-operative University of Kenya.

***Corresponding email:** sandrakatekemboi@gmail.com

Abstract

Each year two hundred and sixty thousand maternal deaths, seventy percent of these deaths happen in Sub-Saharan region. Maternal death continues to be a persistent health issue in Africa despite improvements in maternal healthcare services. Advancements in machine learning technology provide opportunities to predict adverse maternal health outcomes, thereby supporting timely and preventive interventions. While much research has been done to identify the causes of maternal deaths, fewer studies have paid attention to the use of socio-demographic factors in predicting maternal healthcare outcomes. This review aimed to examine evidence on how socio-demographic determinants and machine learning have been applied to predict maternal health outcomes in Sub-Saharan Africa. Electronic databases and scholarly search engines were used to select studies published between January 2019 and June 2026, selected articles included in the review were studies conducted within Sub-Saharan region countries that utilized machine learning algorithms to evaluate how socio-demographic factors affect maternal health outcomes. Frequently used socio-demographic factors associated with maternal healthcare outcomes included household wealth, education, antenatal care utilization, residence place and media exposure. The review noted that use of explainable artificial intelligence such as Shapley Additive exPlanations (SHAP) and association rule mining were used to enhance model interpretability. However, methodological inconsistencies in handling class imbalance and validation procedures were identified as challenges that affect the accuracy of reported models. While socio-demographic determinants provide useful predictive information for maternal healthcare outcomes, particularly in low income settings, future studies should improve methodological consistency and incorporate underrepresented third-delay factors such as quality of care, availability of skilled personnel, medicines, referral efficiency and facility readiness, to support more comprehensive maternal healthcare prediction.

Keywords: Machine learning, maternal health, socio-demographic determinants, maternal healthcare outcomes, Sub-Saharan Africa

Introduction

High maternal deaths continue to be a pressing global public health challenge with a maternal death occurring every two minutes, resulting in about seven hundred deaths daily and an estimated two hundred and sixty thousand maternal deaths occurring globally each year (World Health Organization, 2025). The sub-Saharan region accounts for seventy percent of the global maternal deaths bearing the heaviest burden of approximately one hundred and eighty two thousand maternal deaths yearly (World Health Organization, 2025).

Sub-Saharan countries have, however, made notable progress in reducing maternal deaths with maternal deaths declining by 40% (World Health Organization, 2025). The maternal mortality rate (MMR) decreased from 727 deaths per 100,000 live births in 2000 to 442 per 100,000 live births in 2023. There was also a notable increase in skilled birth attendance, reaching approximately 80%. Across many Sub-Saharan countries, governments have supported the expansion of health infrastructure and subsidize healthcare costs to reduce maternal deaths (Aikpitanyi, 2026). However, maternal healthcare utilization and care-seeking remain influenced by multiple socio-demographic factors that contribute to multidimensional vulnerabilities in accessing and utilizing maternal healthcare services (Aikpitanyi, 2026).

This study is aligned with the established three delays framework proposed by Thaddeus and Maine (1994), which explains maternal deaths in relation to three major delays:

Delay 1; delay in deciding to seek care. This delay is associated with a woman's ability to recognize danger signs, afford healthcare services or make decisions about seeking care independently.



Delay 2; delay in reaching healthcare facility. This delay is associated with distance a woman must travel to reach a health facility, the availability of transport and time required to reach the facility.

Delay 3; delay in receiving adequate care. This delay is associated with ability to receive adequate and timely treatment once at a health facility, which may be influenced by staffing, supplies, and provider readiness.

The advancement and integration of machine learning models in the health sector have opened new possibilities for predicting maternal health-related outcomes (Al Mashrafi et al., 2024). However, existing research and machine learning applications predominantly rely on clinical datasets, which may not fully account for the broader socio-economic, cultural, and demographic factors that are associated with maternal healthcare access and care-seeking decisions in low income countries. This review aimed to identify influential socio-demographic factors incorporated into machine-learning models for predicting maternal health-related outcomes in Sub-Saharan Africa and to examine how machine learning has been applied.

Methodology

The literature search was conducted between February and June 2026 to identify articles published between January 2019 and June 2026, the Search was conducted on four electronic search platforms: Springer Nature Link, Google Scholar, PubMed and IEEE Xplore. Springer Nature Link was used to identify relevant literature published within the Springer Nature collection, specifically BMC journals. These platforms were selected to provide coverage of biomedical, public-health and computer-science literature relevant to machine learning and maternal health. To ensure an effective search, the articles were searched using selected keywords and synonyms which were categorized to represent four categories using Booleans. For Google Scholar, machine-learning-related keywords were restricted to the article title using the intitle: operator. This restriction was applied to improve search precision and manage the large volume of records returned by Google Scholar, the restriction was applied only to Google Scholar and was not applied to PubMed, IEEE Xplore or Springer Nature Link.

Table 1. Search keywords and Boolean operators used to search articles.

Platform	Keyword
Google search	(intitle:"machine learning" OR intitle:"random forest" OR intitle:"gradient boosting" OR intitle:"XGBoost" OR intitle:"SHAP" OR intitle:"deep learning" OR intitle:"explainable artificial intelligence" OR intitle:"predictive modelling" OR intitle:"artificial intelligence") AND ("maternal mortality" OR "maternal death" OR "pregnancy outcomes" OR "pregnancy risk" OR "high-risk pregnancy") AND ("socio-demographic factors" OR "antenatal care" OR "economic status" OR "distance to health facility" OR "education level" OR "social determinants of health") AND ("Sub-Saharan Africa" OR "Kenya" OR "Ethiopia" OR "Nigeria" OR "Uganda" OR "Ghana" OR "East Africa" OR "Africa")
IEEE Xplore	("machine learning" OR "random forest" OR "gradient boosting" OR XGBoost OR "deep learning" OR "predictive modelling" OR SHAP OR "explainable artificial intelligence") AND ("maternal health" OR "maternal mortality" OR "pregnancy risk" OR "high-risk pregnancy" OR "adverse pregnancy outcomes" OR "antenatal care" OR "skilled birth attendance" OR "facility delivery" OR "home delivery" OR "healthcare access") AND ("Sub-Saharan Africa" OR "East Africa" OR Kenya OR Nigeria OR Uganda OR Tanzania OR Ethiopia OR Ghana OR Rwanda OR Zambia OR Malawi OR Somalia)
PubMed	("maternal mortality" OR "maternal death" OR "pregnancy outcomes" OR "pregnancy risk" OR "high-risk pregnancy" OR "maternal health") AND ("socio-demographic factors" OR "antenatal care" OR "economic status" OR "distance to health facility" OR "education level" OR "social determinants of health") AND

	("machine learning" OR "random forest" OR "gradient boosting" OR "XGBoost" OR "SHAP" OR "deep learning" OR "explainable artificial intelligence" OR "predictive modelling" OR "artificial intelligence") AND ("Sub-Saharan Africa" OR "Kenya" OR "Ethiopia" OR "Nigeria" OR "Uganda" OR "Ghana" OR "East Africa" OR "Africa")
Springer Nature Link	("maternal mortality" OR "maternal death" OR "pregnancy outcomes" OR "pregnancy risk" OR "high-risk pregnancy" OR "maternal health") AND ("socio-demographic factors" OR "antenatal care" OR "economic status" OR "distance to health facility" OR "education level" OR "social determinants of health") AND ("machine learning" OR "random forest" OR "gradient boosting" OR "XGBoost" OR "SHAP" OR "deep learning" OR "explainable artificial intelligence" OR "predictive modelling" OR "artificial intelligence") AND ("Sub-Saharan Africa" OR "Kenya" OR "Ethiopia" OR "Nigeria" OR "Uganda" OR "Ghana" OR "East Africa" OR "Africa")

Retrieved articles were exported to Zotero and combined into a single Microsoft excel for deduplication and manual screening to determine if it would be included or excluded. Exclusion and inclusion criteria were developed as the table below shows:

Table 2. Exclusion or inclusion criteria used

Criterion	Inclusion	Exclusion
Geographical Region	Studies conducted in Sub-Saharan African region	Studies conducted outside Sub-Saharan Africa
Methodology	Applied machine learning algorithms methods	Studies that are qualitative or applied only statistical methods
Predictors	Had at least one socio-demographic determinant either wealth index, education, distance to facility, antenatal care attendance, marital status, parity	Used only clinical/ predictors with no socio-demographic variables
Outcome	Predicted adverse maternal health outcomes or related care-seeking outcomes consistent with the Three-Delays framework	Outcome is unrelated to maternal health outcome or pregnancy risk indicator
Publication type	Peer reviewed journal articles, thesis or conference paper	Editorials, commentaries, protocols without primary data
Language	English	Non-English language
Date	Published January 2019 – June 2026	Published before 2019

Initial screening was conducted manually using titles and abstracts followed by full text reading to determine applicability of a study. Two reviewers independently screened the articles against predefined inclusion criteria. Any disagreements about inclusion of a study was discussed and resolved by consensus, where an agreement was not reached, the third reviewer was consulted to make the final decision.

Initial screening had a total of 616 selected articles, after removing 70 duplicate records, 546 records were screened by title and abstract, 532 records were excluded, leaving 14 full-text articles for eligibility assessment. Three articles were subsequently excluded, resulting in 11 studies which were included in the final systematic review. The study followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021).

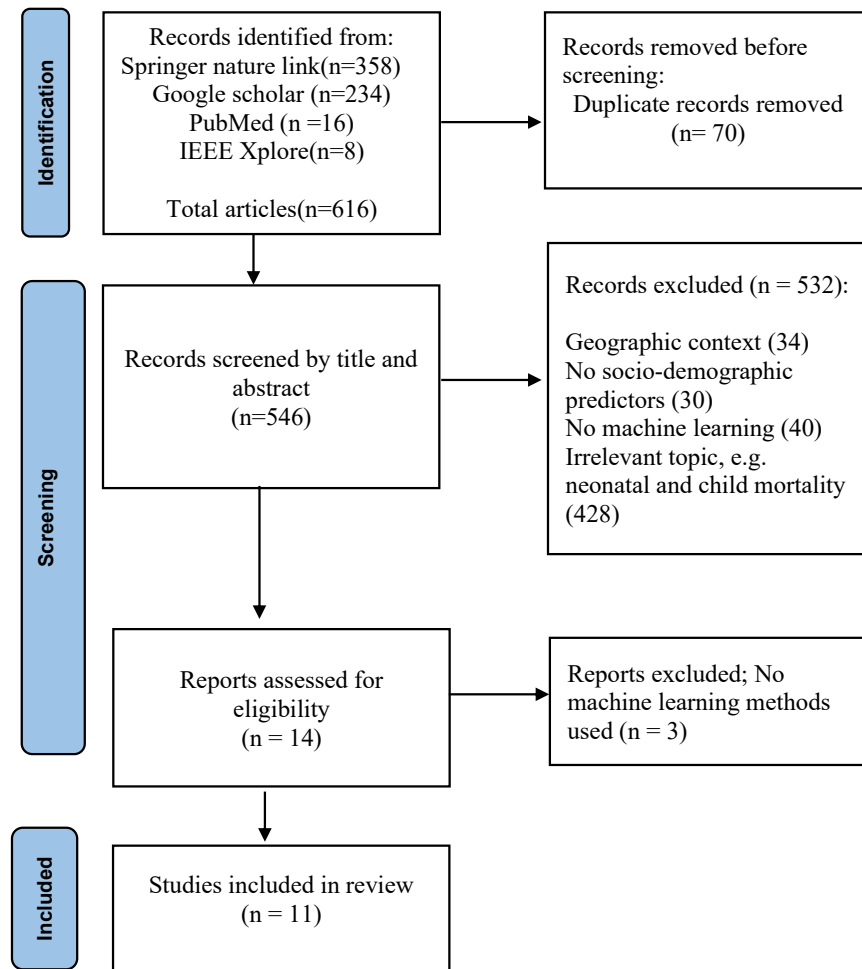


Figure 1. PRISMA flow diagram showing screening and inclusion steps of studies

Risk of bias analysis was assessed independently by both reviewers using Prediction model Risk of Bias Assessment Tool (PROBAST), the 11 selected studies were evaluated across four domains including participants, predictors, outcome and analysis. Each domain was judged as low, high, or unclear risk of bias. A study's overall rating was high if one area was high, unclear one area was unclear and none was high, low rating was given if all four areas were low ;3 studies were rated as low risk, 4 studies as high risk and 4 as unclear. Major concerns were observed in analysis domain which were mainly due to the procedures of balancing in relation to data splitting. Studies were classified as unclear where insufficient methodological information was available to make a definitive judgment.

Data were extracted using a template developed to capture each included study, predicted outcome, socio-demographic predictors used, machine learning algorithms applied, class imbalance and explainability methods as well as alignment with the three delay framework based on the reported predictors and outcomes.

**Table 3. Data extraction and methodological characteristics of included studies**

Author(s),	Sample Size	Outcome Predicted	Socio-demographic Predictors Used	ML Algorithm(s)	Class balancing and explainability methods	ThreeDelays
Sani et al. (2026), Nigeria	16100	Institutional delivery dropout	Age, education, wealth, employment, marital status, religion, residence, region, distance to health facility	Logistic Regression, SVM, KNN, Decision Tree, Random Forest, Gradient Boosting, XGBoost	SHAP for interpretation, Class balance not reported	Delay 1, 2
Osborne et al. (2026), Nigeria	13955	ANC utilization, no, inadequate, adequate	Age, marital status; pregnancy intention, religion, ethnicity, region, residence, media exposure, education, wealth, insurance, employment, healthcare access	LightGBM, XGBoost	SHAP for interpretation, Hierarchical modelling used for class balancing	Delay 1 , 2
Ngaji (2022), Kenya	14898	ANC visits	Timing of first ANC, age, age at first birth, parity, place of delivery, education, radio exposure; wealth, marital status	Random Forest, Logistic Regression	Random Forest feature for interpretation, Class balance not reported	Delay 1
Taye et al. (2025), 27 SSA countries	198707	Skilled birth attendance	Residence, education, internet, wealth, delivery place, marital status, employment, permission, financial capacity, distance, reluctance, media exposure	Random Forest	SHAP for interpretation, SMOTE used for class balancing	Delay 1, 2
Zegeye et al. (2025), Sub-Saharan Africa	299,759	Home delivery	Wealth; mobile phone, ANC visits, education, residence, TV exposure, parity, husband education, marital status, household factors	Random forest, KNN, Decision tree, Linear regression, XGBoost, AdaBoost, ANN, Naive Bayes	SHAP for interpretation, SMOTE used for class balancing	Delay 1,2
Kassaw et al. (2025), rural Ethiopia	885	Health service access	Wealth; residence, media exposure, education	Gradient Boosting, random forest, KNN, Decision tree, SVM, Linear regression, Naive Bayes	SHAP for interpretation, SMOTE used for class balancing	Delay 1,2
Sani et al. (2025), Somalia	8951	Health facility delivery	Wealth, ANC attendance, residence, education, region, marital status, work status, distance, age, births	Random forest, XGBoost, Gradient Boosting, linear regression, SVM, decision tree, KNN	SHAP for interpretation, SMOTE used for class balancing	Delay 1 ,2
Walle et al. (2025), East Africa	44123	Home delivery after ANC	Wealth, rural residence; ANC timing/visits, husband education, contraceptive use, birth interval, distance, media exposure	SVM, Random Forest	SHAP for interpretation, SMOTE used for class balancing	Delay 1, 2
Baykemagn et al. (2025), East Africa	100246	Delayed ANC initiation	Income/wealth, household size, employment, age, family planning, education, residence, home delivery	AdaBoost, DT, Gradient boosting, KNN, LightGBM, Random Forest, XGBoost	Association rule mining; feature importance interpretation, SMOTE used for class balancing	Delay 1, 2



Ngusie et al. (2024), East Africa	86,009	Place of delivery	Working status, media exposure, ANC visits, birth interval, distance, birth order, husband education, household head, age, residence	SVM, CatBoost, XGBoost, random forest, GB, linear regression, decision tree, KNN	association rule mining for interpretation, SMOTE used for class balancing	Delay 1,2
Tesfaye et al. (2019), Ethiopia	11023	Skilled delivery service use	First ANC, birth order, TV ownership, contraceptive use, healthcare cost, age at first birth, age at first sex	J48, Naive Bayes, SVM, ANN, Logistic Regression	Information gain, J48 feature for interpretability, Class balancing not reported	Delay 1

Results and Discussion

Socio-Demographic Predictors of Maternal Health Outcomes

Because study designs, populations and outcome measures were expected to differ no quantitative meta-analysis was planned. The evidence was instead synthesized narratively, following the Synthesis without Meta-analysis reporting guideline (Campbell et al., 2020). Among the 11 included studies, several socio-demographic factors were important in predicting maternal healthcare outcomes. These included education, household wealth, place of residence, distance to health facilities, and media exposure. These factors relate to different aspects of healthcare access, such as knowledge, financial ability, physical access and access to information.

Education was used to predict outcomes such as antenatal care utilization, skilled birth attendance, institutional delivery, and home delivery. Education is important because it improves understanding of health information and awareness of danger signs. For example, a Kenyan study by Ngaji (2022) reported a 23.5% predicted probability of receiving fewer than four Antenatal care (ANC) visits among women with no education. While Sani et al. (2025) also reported that education was associated with hospital delivery, with women with secondary and higher education showing greater likelihood of hospital delivery than women with lower education levels. Within the three delay model, education is relevant to the first delay; education is important in improving health understanding and danger signs therefore prompting the decision to seek healthcare services.

Ngaji (2022) found that timing of the first ANC visit highly influenced the completion of four or more ANC visits. Each month's delay in starting ANC showed lower odds of completing the recommended four or more visits (OR = 0.233, 95% CI 0.220–0.246). The same study indicated that women with previous births were more likely to complete four or more clinical visits with each additional child associated with an approximately 9.7% increase in the odds of completing four or more visits (OR = 1.097, 95% CI 1.068–1.127). Decision to seek antenatal care is associated with the first or second delay leading to maternal deaths as the delay might be because of financial constraints or the distance to health facilities.

Household wealth was one of the common socio-demographic factors identified, wealth affects a woman's ability to meet healthcare-related costs, including transportation and treatment expenses. In the reviewed studies, lower household wealth was associated with low access to maternal healthcare services. For example, Baykemagn et al. (2025) identified wealth as a relevant predictor of delayed antenatal care initiation affecting the number of ANC visits a woman will attend to, wealth is a strong indicator associated with the first delay.

Media exposure, an indicator associated with the first delay, was identified as a relevant socio-demographic. Media such as television, telephones or radio provide women with access to information about maternal healthcare services. Baykemagn et al. (2025), with a sample of 77,865 women from 10 East African countries, reported that 67.3% of women without mobile phones had delayed antenatal care initiation, compared with 61.8% of women who owned mobile phones. Similarly, Tesfaye et al. (2019) findings show that television ownership was an important predictor of skilled delivery among women in Ethiopia. Another study by Zegeye et al. (2025) supported the claim of using television and mobile phone ownership as predictors of maternal healthcare use among women in Sub-Saharan.

Place of residence and distance to a health facility determined the level of maternal healthcare utilization. These factors represent the geographical aspect of access to maternal healthcare, representing geographical dimension of access to

maternal healthcare related to the second delay. Sani et al. (2025) did a study in Somalia that showed 34.7%, 29.9% and 35.3% of the women lived in urban, rural, and nomadic settings, respectively. Residence area was associated with health-facility delivery because women in urban and rural had lower odds of delivering in a health facility (AOR = 0.71, 95% CI: 0.59–0.86), while nomadic women had substantially lower odds (AOR = 0.23, 95% CI: 0.16–0.32). Similarly, Taye et al. (2025) also identified difficulty in accessing health facilities as a relevant predictor of home delivery. The study reported 71% of women received skilled birth attendance, a further SHAP analysis identified urban residence and shorter distance to healthcare facilities among predictors contributing to skilled birth attendance.

Machine learning Approaches and Methodology

Twelve distinct machine learning algorithms were identified across the eleven studies including; Random Forest, Support vector machine, Logistic regression, Decision tree, Gaussian naive bayes, K-nearest neighbours, Gradient boosting, XGBoost, LightGBM, CatBoost, AdaBoost, and Artificial neural networks. Several studies evaluated multiple algorithms, in modelling approaches demonstrating considerable variation across literature, the varied algorithm selection across studies shows model algorithm selection depended on modelling objectives, datasets and characteristics of the prediction tasks.

Class imbalance was an important methodological consideration across the included studies. SMOTE or a SMOTE variant was used in nine articles to address imbalanced outcome classes. However, reporting of the sequence of class balancing in relation to data splitting was inconsistent. Walle et al. (2025) explicitly reported performing data balancing on the training data after the train-test split step, while Taye et al. (2025) and Tesfaye et al. (2019) described balancing before splitting, the remaining studies does not clearly report the order of these procedures. This finding draws attention to the importance of clearly reporting the timing of class-balancing procedures, as this is relevant to assessing the reliability and generalizability of the model.

SHAP was used as an interpretability method in several of the included studies, other notable approaches included association rule mining and model-based feature importance. These approaches were used to interpret model outputs and identify predictors that contributed to model predictions.

Limitations of the study

It is important to note that the study has some limitations such as;

1. Only studies in English were included in the review, studies published in other languages may have been missed because of this criterion.
2. The use of the intitle restriction in Google Scholar may also have excluded studies that used machine learning but did not mention it in their titles.
3. Studies included in the review had different maternal health outcomes such as ANC utilization, delayed ANC initiation, skilled birth attendance, institutional delivery, home delivery and healthcare access. These studies also used different machine-learning algorithms and evaluation measures therefore a direct comparison of model performance was difficult.

Conclusion and Recommendations

The review indicates that socio-demographic indicators play a crucial role in prediction of maternal healthcare in Sub-Saharan Africa. Household wealth, education, media exposure and distance to health facilities emerged as relevant predictive variables across the reviewed studies, highlighting the importance of considering socioeconomic and geographic dimensions when developing predictive models for maternal healthcare outcomes.

Media exposure emerged as a relevant predictive variable in several of the reviewed studies, suggesting that access to health information through media channels contributes to maternal healthcare utilization, therefore, health interventions and education may consider diverse media channels and locally appropriate languages to improve healthcare utilization, particularly in rural communities.



Geographic accessibility plays a role in maternal healthcare utilization, in areas where there is difficulties in getting to the hospital interventions such as mobile clinics may be considered as potential strategies for addressing geographic barriers to maternal healthcare.

There is limited incorporation of third-delay factors such as quality of care, availability of skilled personnel/ medicines and facility readiness, into existing predictive models. Future research should therefore examine how these factors can be incorporated to inform maternal health in Sub-Saharan Africa.

References

- Aikpitanyi, J. (2026). Access alone is not enough: Rethinking maternal health in sub-Saharan Africa. *African Journal of Reproductive Health*, 30(3), 9–12. <https://doi.org/10.29063/ajrh2026/v30i3.1>
- Al Mashrafi, S. S., Tafakori, L., & Abdollahian, M. (2024). Predicting maternal risk level using machine learning models. *BMC Pregnancy and Childbirth*, 24, Article 820.
- Baykemagn, N. D., Taye, E. A., Melaku, M. S., Yehuala, T. Z., Tewelgne, M. F., Arage, F. G., & Zegeye, A. T. (2025). Predicting delayed antenatal care initiation among pregnant women in East Africa: Using machine learning algorithms. *Frontiers in Global Women's Health*, 6, 1488391.
- Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, 368, l6890. <https://doi.org/10.1136/bmj.l6890>
- Kassaw, A. K., Kidie, A. A., Tesfa, N. A., Mislu, E., Yimer, A., Hailu, M., Yesuf, H. A., Ayal, B. G., Geto, A. K., & Feleke, S. F. (2025). Optimizing machine learning models for predicting health service access and determinants among pregnant women in rural Ethiopia. *Scientific Reports*, 15(1), 40559.
- Ngaji, C. O. (2022). Predicting antenatal care (ANC) visits using machine learning algorithms [Master's thesis, University of Nairobi].
- Ngusie, H. S., Tesfa, G. A., Taddese, A. A., Enyew, E. B., Alene, T. D., Abebe, G. K., Walle, A. D., & Zemariam, A. B. (2024). Predicting place of delivery choice among childbearing women in East Africa: A comparative analysis of advanced machine learning techniques. *Frontiers in Public Health*, 12, 1439320.
- Taye, E. A., Woubet, E. Y., Hailie, G. Y., Arage, F. G., Zerihun, T. E., Zegeye, A. T., Zeleke, T. C., & Kassaw, A. T. (2025). Application of the random forest algorithm to predict skilled birth attendance and identify determinants among reproductive-age women in 27 Sub-Saharan African countries; machine learning analysis. *BMC Public Health*, 25(1), 901.
- Tesfaye, B., Atique, S., Azim, T., & Kebede, M. M. (2019). Predicting skilled delivery service use in Ethiopia: Dual application of logistic regression and machine learning algorithms. *BMC Medical Informatics and Decision Making*, 19(1), 209.
- Thaddeus, S., & Maine, D. (1994). Too far to walk: Maternal mortality in context. *Social Science & Medicine*, 38(8), 1091–1110.
- World Health Organization. (2025). Trends in maternal mortality estimates 2000 to 2023: Estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division
- Zegeye, A. T., Tilahun, B. C., Fekadie Tewelgne, M., Addisu, E., Wassie, B., Alelign, B., Sharew, M., Baykemagn, N. D., Kebede, A., & Yehuala, T. Z. (2025). Predicting home delivery and identifying its determinants among women aged 15–49 years in sub-Saharan African countries using a Demographic and Health Surveys 2016–2023: A machine learning algorithm. *BMC Public Health*, 25(1), 302.