DOI: https://doi.org/10.61762/pjstvol9iss1art003

PLATFORM - A Journal of Science & Technology

Received: 23 February 2026, Revised: 10 April 2026, Accepted: 28 April 2026, Published: 30 June 2026, Publisher: UTP Press, Creative Commons: CC BY 4.0

PREDICTING MATERNAL HEALTH OUTCOMES USING DECISION TREE MODELS: RESULTS AND LIMITATIONS FROM RURAL SUDAN

Rayan Rahmtallah[1]*, Ashraf Osman Ibrahim[2], Mohammed Merghany Mohammed Abdelsalam[1]

[1]Faculty of Computer Science and Information Technology, Alzaiem Alazhari University, Sudan
[2]Department of Computing, Universiti Teknologi PETRONAS, Perak, Malaysia

*E-mail: rayanwaheibkhalifa@gmail.com

ABSTRACT

This study addresses the critical challenge of maternal health resource misallocation in rural Sudan, which significantly impacts maternal mortality rates. Leveraging data mining techniques, specifically the Decision Tree algorithm, pregnancy care data (n=2,201) collected by the Ministry of Cabinet Central Bureau of Statistics were analysed. The primary objective was to develop a predictive model to classify the outcomes of the last completed pregnancy (Alive, Dead, or Abortion) to support evidence-based decision-making. Utilising a hold-out test set of 331 instances, the model achieved a 100% classification accuracy, perfectly identifying outcomes for 293 'Alive', 30 'Dead', and 8 'Abortion' cases. The findings highlight key predictors, including geographic location, access to professional antenatal care (ANC), and the importance of routine diagnostic tests such as blood sampling. The study provides actionable insights for policymakers to optimise resource allocation and enhance maternal survival through targeted interventions in high-risk rural areas.

Keywords: *Maternal health, data mining, decision tree, random forest, decision tree models, health informatics, machine learning*

INTRODUCTION

Despite significant global progress in maternal health, maternal mortality remains unacceptably high in many low- and middle-income countries, particularly in Sudan. Access to essential maternal health services, including antenatal care (ANC), skilled birth attendance, and postnatal care, continues to be limited, especially in rural areas where geographic barriers, shortages of skilled healthcare providers, and unequal distribution of health resources persist. These structural challenges contribute to poor maternal health outcomes and highlight the need for more effective, data-driven approaches to support decision-making.

In Sudan, the situation is further complicated by spatial inequalities in healthcare access between urban centres such as Khartoum and rural or underserved regions. Women in rural areas face increased risks during pregnancy due to delayed or insufficient access to care. Although previous studies have examined determinants of maternal health service utilisation using conventional statistical methods, these approaches are often limited in their ability to capture complex, non-linear relationships among multiple influencing factors. As a result, important hidden patterns within large health datasets may remain unexplored.

Recent advancements in data mining provide powerful tools for analysing large and complex datasets, enabling the identification of patterns and relationships that can improve predictive accuracy and inform targeted interventions. However, a gap persists in the effective application of these techniques in maternal health in Sudan, particularly in rural settings. Furthermore, limited research has focused on predicting specific maternal outcomes, such as the outcome of the last completed pregnancy, using robust, data-driven models.

This study aims to address these gaps by applying data mining techniques to maternal healthcare data from rural Sudan to predict the outcome of the last completed pregnancy (alive, dead, or abortion). Specifically, the study develops a decision tree–based predictive model to classify maternal health outcomes and identify key risk factors associated with adverse outcomes. By doing so, this research contributes to the existing literature by demonstrating the applicability of predictive analytics in maternal health and providing an interpretable model that can support evidence-based decision-making.

The findings of this study are expected to assist policymakers, healthcare providers, and development organisations in improving resource allocation, identifying high-risk cases, and prioritising interventions in resource-constrained settings. Ultimately, this work aims to help reduce maternal mortality and improve maternal health outcomes by leveraging data-driven insights to inform more effective and equitable healthcare strategies.

LITERATURE REVIEW

In 2002, Adamu et al. [1] investigated barriers to the use of ANC services and hospital delivery in rural Kano State, Nigeria. The study identified socio-cultural and economic constraints influencing women's utilisation of ANC services. The findings showed that financial limitation (46%), predestination beliefs (17.2%), and husband's refusal (17.2%) were the major barriers to service utilisation. The study highlighted the importance of understanding behavioural and socioeconomic determinants when modelling maternal health service utilisation.

In 2015, Mugo et al. [2] analysed the prevalence and determinants of non-use of ANC services in South Sudan using data from the South Sudan Household Survey. The study reported that 58% of women did not attend ANC visits, while only 18% completed the recommended four visits. Geographic region, literacy level, polygamy status, and knowledge of newborn danger signs were significant predictors of ANC non-utilisation. The study emphasised the need for predictive approaches to identify high-risk groups for targeted intervention.

Nzioki et al. [3] examined socio-demographic factors influencing maternal and child health (MCH) service utilisation in Mwingi District, Kenya. Using logistic regression analysis on data collected from 416 women, the study found that age, education level, poverty status, and geographical area significantly influenced ANC attendance. The authors recommended predictive modelling approaches to assist policymakers in improving maternal health utilisation.

In 2016, Sahle [4] applied Knowledge Discovery in Databases (KDD) techniques to identify determinants hindering postnatal care (PNC) visits in Ethiopia. The study employed J48 Decision Tree and JRip rule induction algorithms on 6,558 EDHS records. The J48 model achieved an accuracy of 93.97%, outperforming JRip slightly. Place of delivery, assistance from skilled professionals, attendance at prenatal care, and maternal age were identified as strong predictors of PNC utilisation. This study demonstrated the effectiveness of Decision Tree models in predicting maternal service utilisation.

Cananua-Labid [5] conducted a CHAID (Chi-square Automatic Interaction Detector) analysis using data from 5,301 women obtained from the Philippine National Demographic and Health Survey in 2017. The CHAID decision tree model identified region, religion, wealth quintile, and educational level as key determinants of ANC utilisation. The model successfully segmented vulnerable groups of women who require policy attention, demonstrating the applicability of tree-based models for utilisation prediction.

In 2018, Abegaz [6] developed a predictive model of skilled birth attendant utilisation among ANC clients in Ethiopia using EDHS 2016 data. The study applied J48 Decision Tree, PART, and Naïve Bayes algorithms within WEKA. The J48 model achieved an accuracy of 98.74%. The most influential predictors were residence, region, husband's education level, wealth index, and media access. The study confirmed that decision tree models are highly effective at predicting skilled birth attendant utilisation.

Abegaz et al. [7] explored trends and barriers to ANC utilisation in Ethiopia using classification, clustering, and association rule mining techniques on 28,631 EDHS records (2000–2016). The study identified education level, residence, media exposure, and economic status as consistent predictors of ANC service utilisation. The authors emphasised that predictive modelling can support targeted maternal health interventions.

A master's thesis submitted to Addis Ababa University developed a predictive model entitled "Predicting Maternal Health Care Seeking Pattern Using Data Mining Techniques in Ethiopia." The study applied J48 Decision Tree and Naïve Bayes algorithms to 16,515 records covering antenatal, delivery, and postnatal periods. The model achieved accuracies of 74.78% for ANC, 91.23% for delivery care, and 87.5% for postnatal care. The research demonstrated that decision tree algorithms are practical tools for predicting maternal health care utilisation patterns [6].

Recently, Sanni et al. [8] applied Random Forest and Decision Tree models to predict optimal ANC utilisation among Nigerian women, using data from the Demographic and Health Survey. The study showed that tree-based models outperformed traditional regression models in identifying women at risk of inadequate ANC attendance, highlighting wealth index, education level, and rural residence as major predictors.

Mlandu, Matsena-Zingoni, and Musenge (2024) applied decision tree models to predict the non-utilisation of postnatal care services using Demographic and Health Survey data from three East African countries. The study demonstrated that decision tree algorithms can effectively identify complex relationships among maternal and healthcare factors influencing postnatal care utilization and provide interpretable decision rules for healthcare planning and intervention design [9].

Taye et al. applied a Random Forest algorithm to predict skilled birth attendance among reproductive-age women in 27 Sub-Saharan African countries. The study demonstrated strong predictive performance, achieving 92% accuracy and identifying important determinants including maternal education, household wealth, urban residence, healthcare accessibility, media exposure, and internet use. The findings highlighted the potential of machine learning techniques to support maternal health planning and targeted interventions [10].

RESEARCH METHODOLOGY

The proposed framework is structured around a systematic KDD pipeline, specifically tailored to handle the complexities of maternal health surveys in Sudan. The framework consists of five integrated phases:

1. Data Selection and Identification: In this phase, domain-specific attributes were identified through a comparative analysis between the Sudan MICS dataset and existing maternal health studies. This ensured that the model focuses on the most clinically relevant variables for the Sudanese context.
2. Data Preprocessing and Imputation: To address missing values, a 'Mode Imputation' strategy was applied to categorical variables. This step is critical for maintaining the dataset's statistical power while preventing the bias that often results from simply excluding incomplete records.
3. Statistical Validation (Feature Selection): A Chi-Square test of independence was utilised as a filter method to validate the relationship between each predictor and the maternal outcome. Only variables with a p-value < 0.05 were retained, ensuring a lean and efficient model.
4. Model Development (Decision Tree Induction): The core of the framework utilises the Decision Tree algorithm. This choice was dictated by the model's inherent 'interpretability,' which is essential for healthcare practitioners and policymakers who need to understand the logic behind each prediction.
5. Evaluation and Interpretation: The framework concludes with a rigorous evaluation phase using a hold-out test set and a confusion matrix to verify the model's reliability and its ability to generalise to new, unseen health data.

Research Design

This study adopts a quantitative data mining approach based on the KDD process, which comprises data selection, data preprocessing, data transformation, data mining, and model evaluation. The KDD framework provides a systematic methodology for extracting meaningful patterns and predictive knowledge from large datasets while ensuring methodological rigor, reliability, and transparency throughout the research process [11]-[12].

Data Preparation and Feature Engineering

Data were obtained from the Sudan Multiple Indicator Cluster Survey (MICS), developed by UNICEF and the Central Bureau of Statistics [13]. The Women's Health module was selected due to its relevance to maternal and newborn health, consistent with prior studies [14]. Feature selection was informed by findings from previous maternal health studies conducted in Ethiopia and Sudan, which identified key socio-demographic, healthcare utilization, and pregnancy-related factors associated with maternal health outcomes [15]-[16].

Data Preprocessing

Preprocessing involved removing irrelevant attributes, harmonising variable names, correcting inconsistencies, and standardising formats. Python and Jupyter Notebook were used for data exploration and preprocessing [17]. Missing values were handled using mode imputation, a common approach in categorical health datasets [18]-[19].

The preprocessing phase involved:

1. Removal of irrelevant and duplicate attributes
2. Harmonisation of variable names across survey modules
3. Correction of inconsistent or erroneous entries
4. Standardisation of attribute formats

Python programming language and Jupyter Notebook were used for data exploration, preprocessing, transformation, mining, and visualisation [13].
Attributes with equivalent meanings but different variable names were harmonised. For example:

1. *"Health care check-ups during this pregnancy"* was mapped to ANC
2. *"Outcome of completed pregnancy"* and *"Total children ever born"* were standardised

Question-based variable names were rephrased into descriptive attribute names to improve interpretability, as recommended in health data preprocessing literature [14].

Handling Missing Values

Missing data is a persistent challenge in large-scale health datasets. Removing records with missing values may introduce bias, reduce statistical power, and result in substantial information loss [15]. Therefore, imputation strategies were adopted rather than listwise deletion. Given that all selected attributes were categorical, missing values were handled using mode imputation, whereby missing entries were replaced with the most frequently occurring value for each attribute. Mode imputation is a widely accepted preprocessing technique for categorical datasets because it preserves data distribution and improves dataset completeness prior to model development [12]. This approach has also been applied in healthcare and maternal health data mining studies to enhance data quality and predictive performance [16].

Data Transformation

Data transformation was performed to prepare the dataset for machine learning modelling. This process included:

1. Encoding categorical variables into numerical formats
2. Normalisation and restructuring of attributes
3. Role assignment for predictor and target variables

Label encoding was applied to ordinal categorical variables, while binary encoding was used for nominal attributes. Statistical association tests, including chi-square analysis, were conducted to evaluate relationships between predictors and the target variable, as recommended for categorical health data analysis [17]-[18].

Data Exploration and Correlation Analysis

Exploratory data analysis was conducted to identify relationships between predictor variables and the target label, *Outcome of Last Completed Pregnancy*. Visualisation techniques, including correlation matrices and heat maps, were used to identify highly correlated features [20].

Initial analysis revealed a high correlation between geographic variables (state and residence), suggesting that association rule mining alone would be insufficient. Consequently, a decision tree approach was selected, consistent with prior healthcare modelling studies [18],[21].

A custom correlation function was implemented in Python to identify features exceeding a predefined correlation threshold. Six features with low correlation were removed to reduce dimensionality and avoid redundancy, following best practices in feature selection and feature reduction [22].

Figure 1 illustrates correlations among the main predictors, while Table 1 summarises feature dependencies and correlation values, highlighting the features with the highest correlation in this research dataset.

Chi-Square Analysis for Categorical Features

Chi-square tests were applied to evaluate associations between categorical predictors and the outcome variable. Separate analyses were conducted for ANC, assistance at delivery, and postpartum assistance. All chi-square tests yielded p-values below 0.05, indicating statistically significant relationships between these variables and pregnancy outcomes. These findings are consistent with previous maternal health utilisation studies [14]. Consequently, all related categorical features were retained for model development.

The feature engineering phase was critical in transforming raw survey data into a high-quality input for the decision tree model. By aligning common features from multiple studies in Sudan and Ethiopia, we identified ten core attributes that represent the most significant determinants of maternal health. To validate these selections, a Chi-Square test of independence was performed. The results demonstrated that all selected categorical predictors, including ANC, assistance at delivery, and blood sampling, were statistically significantly associated with the pregnancy outcome ($p < 0.05$).

The core attributes for this study were identified through a comparative analysis of established maternal health research in Sudan and Ethiopia to ensure clinical relevance within the local context. Table 2 presents a detailed comparison of these common features across studies from Ahfad University, Addis Ababa, and Alzaiem Alazhari University, providing the foundation for the final feature selection.

Based on this comparative framework, the selected dataset originated from the Sudan Household Health Survey (SHHS/MICS 2006), specifically focusing on the women's health module and maternal health indicators. Subsequently, the correlation analysis revealed a strong correlation among clinical indicators, including blood, urine, and iron tests ($r \approx 0.87$), prompting the model to prioritise these as key decision nodes. This rigorous filtering process not only reduced the dimensionality of the dataset by removing six redundant features but also ensured that the resulting decision tree is grounded in both clinical evidence and statistical significance.

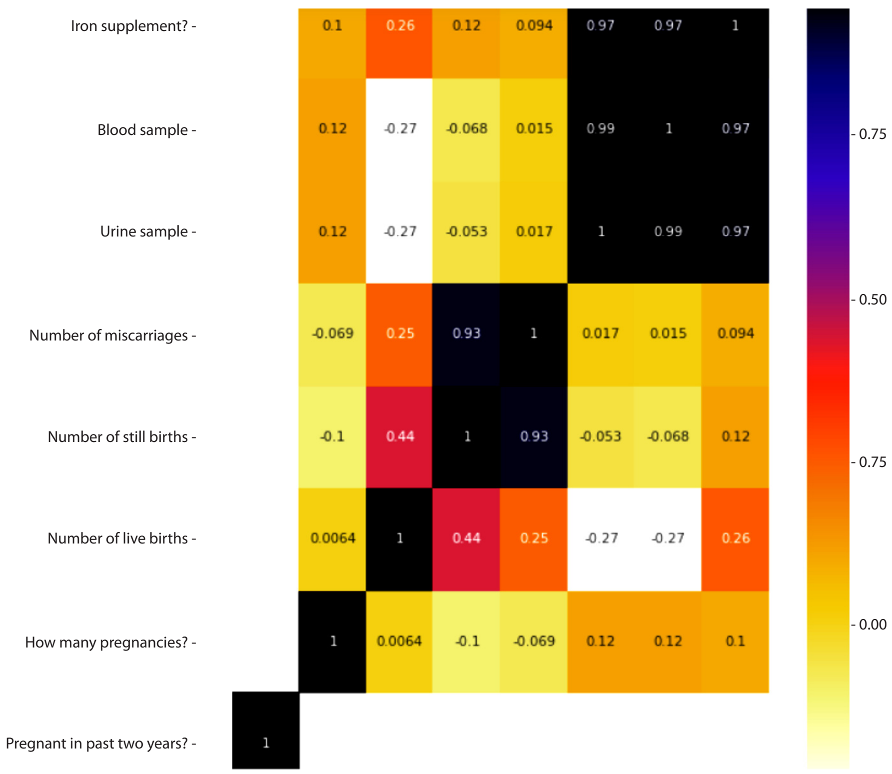

Figure 1 shows the high correlation between the main predictors and the other attributes

Table 1 shows the correlation and dependency of the main feature

Feature categories	High correlation	Correlation value/P-value	Highest Independent feature
Bio-obstruct	Pregnant in the last two years	0.58	Number of stillbirths
Women during pregnancy	Blood sample Urine sample Iron supplement? Bleeding	0.87 0.87 0.56 0.57	Blood sample
Risk during delivery	Hypertension Edema Headache Fever Abdominal pain Convulsions Urinary pain Jaundice Severe breathlessness	0.81 0.81 0.84 0.87 0.85 0.82 0.84 0.82 0.83	Abdominal pain
Mode of delivery	Mode of delivery	0.59	Natural birth, Birth by vacuum or pump, Caesarean section
Place of delivery	Place of delivery	0.58	Home, Primary Healthcare Centre, Primary Healthcare Unit/Hospital, Government Hospital, Private Hospital, Other.
ANC	Doctor, nurse/midwife, auxiliary midwife, traditional birth attendant, community health worker, relative/ friend, no one	0.000000e+00, 0.000000e+ 00, 0.000000e+00,0.000000e+ 00, 2.292037e-35, 2.077128e-308, 1.510956e-105, 0.000000e+00	doctor, nurse/midwife, auxiliary midwife, traditional birth attendance, relative/ friend, no one
Assistant at delivery	Doctor, nurse/midwife, auxiliary midwife, traditional birth attendant, community health worker, relative/ friend, no one	2.322689e-300, 0.000000e+00, 0.000000e+00, 0.000000e+00, 1.121915e-14, 8.201304e-211, 7.938913e-117, 0.000000e+00	doctor, nurse/midwife, auxiliary midwife, traditional birth attendance, community health worker, no one

Model Design and Development

A decision tree classification model was developed using a 10-fold cross-validation strategy to enhance reliability and reduce overfitting [23]. The dataset was randomly divided into training (85%) and validation (15%) subsets. Decision trees were selected for their ability to handle mixed data types and to generate interpretable decision rules suitable for maternal health policy and decision-support contexts [16]-[17].

Model Building

The model was implemented in Python using Jupyter Notebook, following a standard machine learning workflow [14],[20]:

1. Importing required libraries
2. Loading and exploring the dataset
3. Preparing feature and target variables
4. Splitting data into training and testing sets
5. Training the decision tree model
6. Evaluating model performance

Model Results

The confusion matrix Table 4 demonstrates model performance across three categories: alive, dead, and abortion. Out of 331 test instances, the model achieved 100% classification accuracy, with no false positives or false negatives. While this indicates strong internal performance, similar studies caution that such results may reflect data-specific characteristics [19],[23]. Therefore, further external validation with independent, more recent datasets is recommended to confirm generalisability and robustness [6]. The confusion matrix Table 4 demonstrates model performance across three categories: alive, dead, and abortion.

Methodological Strengths and Applicability

The proposed framework demonstrates several methodological strengths. It provides a transparent,

Table 2 Comparison of the common features between the Ethiopian studies, the Addis Ababa study, and the Ahfad University study

Addis Ababa study (17 attributes)	Ethiopic maternal health study (14 attributes)	Ahfad University study (13 attributes)	Alziem Alazhari study (10 attributes)
Antenatal care	Antenatal care	Use pregnancy care	Health care check-ups during this pregnancy
Region	Region	Region	-
Residence	Residence	Residence	-
Mother age	Age	Age	-
Women's Level of education	Women's Level of education	Women's Level of education	-
Husbands occupation	Place of delivery	Our favourite place for Birth?	Your location of delivery?
Religious	Religious	Why do women prefer home birth?	Type of birth
Postnatal care	Postnatal care	Did a health team ever come to your area?	Have you seen anyone for health care during this pregnancy?
Marital status	Ahealthpds	Married status	-
Total children ever born	Pprofessional	-	Outcome of a completed pregnancy
Assistance during Delivery	Pother	What is your opinion on Pregnancy?	Have you seen anyone for health care during this pregnancy?
Frequency of reading newspaper	Dlbycaesariansction	Information source	Information or advice about AIDS or HIV
Frequency of Watching TV	Tetanus injection	The best way to inform people	Blood pressure test
Respondents occupation	-	-	Urine test
Household wealth index	-	-	Blood test
Husband Level of education	-	-	Pregnancy complications
Frequency of listening to the radio	-	-	-

Table 3 Conceptual overlap of variables identified across the four maternal health studies

	Attributes common across all studies.
	Attributes shared between Ethiopian maternal health, Addis Ababa University, and Ahfad University studies.
	Ethiopic maternal health attributes, Ahfad University study, and Alziem Alazhari studies.
	Addis Ababa study and Ethiopic maternal health study.
	Ahfad University attributes and Alziem Alazhari studies.
	Attributes shared between Addis Ababa University studies and Ahfad University studies.
	Attributes Addis Ababa study and Alziem Alazhari studies.

Table 4 Confusion matrix of the decision tree

Class	(Alive)	(Dead)	(Abortion)
Predicted Outcome of last completed pregnancy (Alive)	293	0	0
Predicted Outcome of last completed pregnancy (Dead)	0	30	0
Predicted Outcome of the last completed pregnancy (Abortion)	0	0	8

step-by-step approach to maternal health data mining that can be replicated in other low-resource settings. The use of decision trees enhances interpretability, enabling health practitioners and policymakers to translate analytical outputs into actionable insights. The conceptual comparison showed the intersection among variables, as shown in Table 3.

Furthermore, aligning the workflow with the KDD process ensures methodological consistency and facilitates comparisons with future studies that use alternative machine learning techniques.

The Decision Tree Mining Algorithm

The decision tree used to predict the outcome of the last completed pregnancy, its origin (state), knowledge about pregnancy care, place of delivery (including home birth), the mode of delivery, and ANC and assistance at delivery. Researchers and experts believe that factors such as Sudanese culture, environment, and improper data collection and entry affect feature selection. E.g. marital status is not needed because our assumption about women who did the pregnancy care is that they are married by default, whereas in other cultures, single women do the pregnancy care. Another assumption is that all women receiving pregnancy care are of reproductive age, so there is no need for women's age attributes in mining. The analysis will focus on women during pregnancy and their states, risks during delivery, mode of delivery, place of delivery, ANC, assistance at delivery, and outcome of the last completed pregnancy.

The researcher found that in several Sudanese states, particularly Jonglei, Central Equatoria, Western Bahr el Ghazal, and Warrap, the majority of women reported a higher number of deaths from their last completed pregnancies compared to other regions. In contrast, states such as Sennar, Khartoum, Kassala, Gedarif, River Nile, and Western Equatoria recorded higher numbers of live births and fewer pregnancy-related deaths. Additionally, cases of abortion were mainly observed in Northern and Warrap states, while no abortions were reported in White Nile, Kassala, Western Equatoria, Northern Bahr el Ghazal, and Eastern Equatoria, as shown in Figure 2.

Secondly, women who receive a blood sample test during pregnancy care are expected to have better outcomes in their last completed pregnancy than women who miss the blood sample test during pregnancy care. Despite 95, 50% of women knowing the importance of a blood sample during pregnancy care, they are expected to have an outcome of the

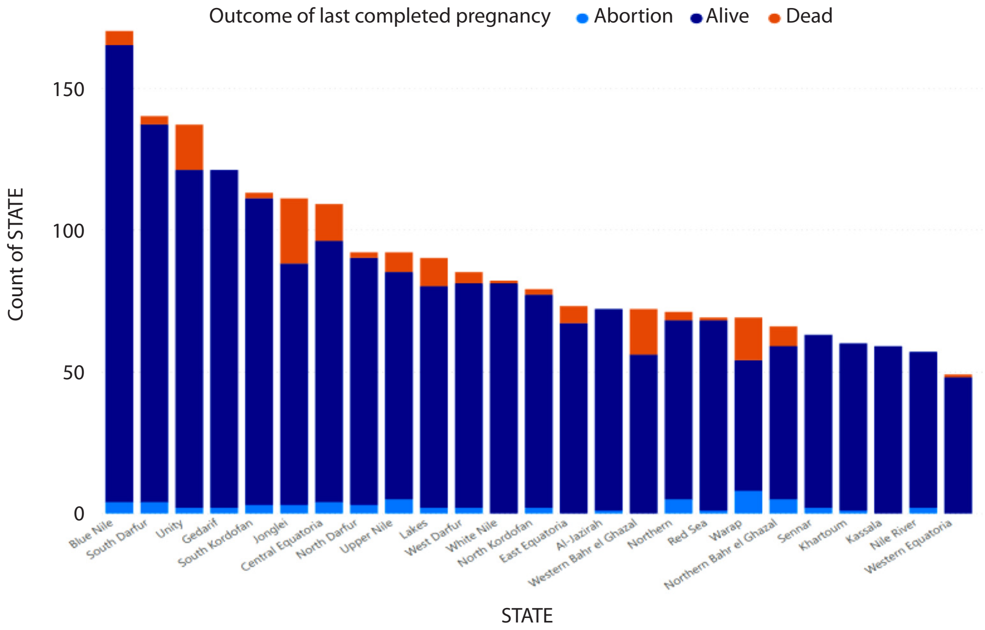

Figure 2 Sudan States and the outcome of the last completed pregnancy of women who were alive, dead or aborted

last completed pregnancy alive. There are still 3.75% of women who did not do the blood sample during pregnancy care, and the expected outcome of the last completed pregnancy is dead. So, the focus of decision-makers will be on obtaining a blood sample during pregnancy to avoid the outcomes of the last completed pregnancy.

Figure 3 shows the distribution of pregnancy outcomes according to blood sample collection during pregnancy. The majority of women who did not have a blood sample taken experienced a live birth, followed by a smaller number of death and abortion outcomes. Similarly, live births were the most common outcome among women who received blood sample testing. Overall, the figure indicates that live birth was the predominant pregnancy outcome across all categories of blood sample status.

The knowledge discovered from the Decision Tree model is illustrated in Figure 4. The model identified several patterns associated with pregnancy outcomes and maternal healthcare utilisation. The findings suggest that pregnancy outcomes varied according to the type of healthcare provider consulted during antenatal care (ANC) and delivery. In particular, adverse outcomes, including death and abortion, were associated with limited utilisation of skilled maternal healthcare services. The analysis further revealed that many women relied on relatives, friends, and traditional birth attendants for ANC more frequently than they consulted doctors or other skilled healthcare providers. This pattern may reflect challenges related to access, affordability, availability of healthcare services, or cultural preferences in rural areas. The findings highlight the need to strengthen access to skilled maternal healthcare services and promote the utilisation of qualified healthcare professionals during

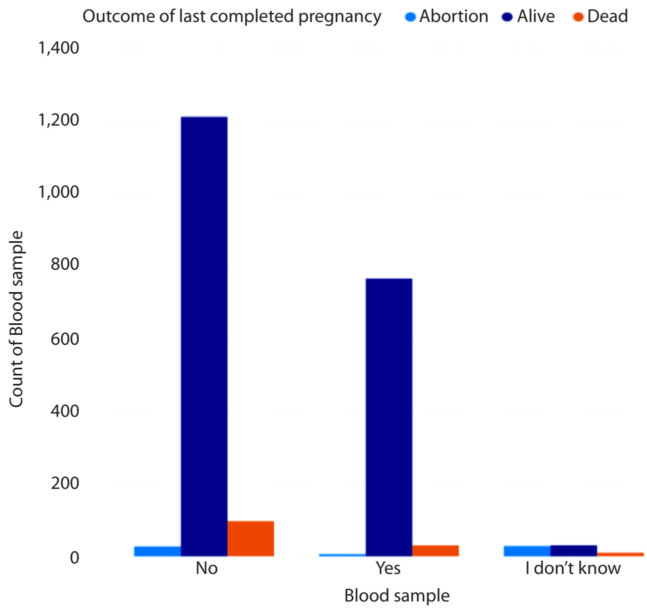

Figure 3 Outcome of the last completed pregnancy and blood sample test

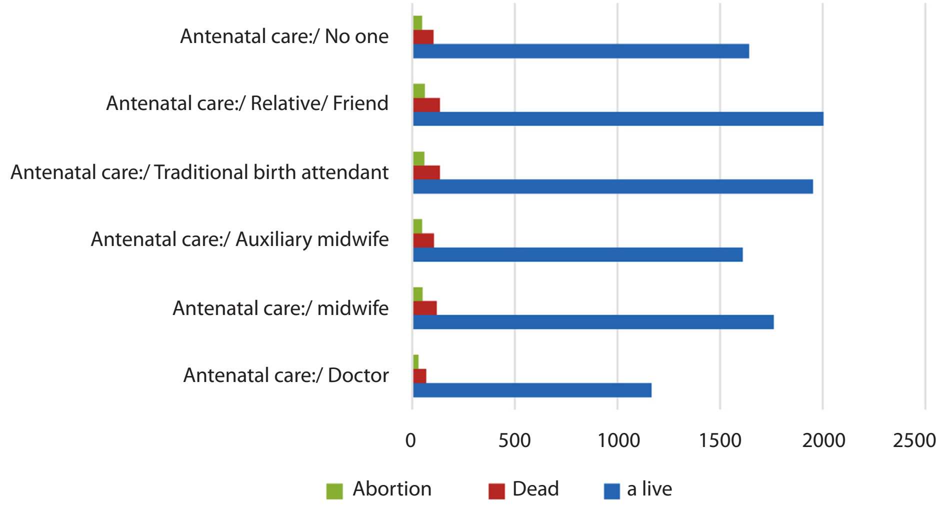

Figure 4 The relationship between ANC and the outcome of the last completed pregnancy

pregnancy and childbirth to improve maternal health outcomes in Sudan.

The results of the Decision Tree analysis revealed that the majority of last completed pregnancies resulted in live births (Figure 5). The model indicated that women who received delivery assistance from community health workers or traditional birth attendants were more frequently associated with live birth outcomes than those who relied solely on doctors or had no professional assistance during delivery. Furthermore, the analysis suggested differences in pregnancy outcomes according to the type of birth attendant, with women assisted by nurses/midwives and auxiliary midwives exhibiting distinct outcome patterns. These findings highlight the importance of skilled and accessible maternal healthcare services in influencing pregnancy outcomes and underscore the need for further investigation into the role of different healthcare providers in improving maternal health in Sudan.

Figure 6 illustrates the distribution of pregnancy outcomes according to the mode of delivery. The figure shows that natural birth was the most common mode of delivery and accounted for the highest number of live births among women in the study. However, natural birth was also associated with a higher number of death and abortion outcomes compared to the other delivery methods, largely because it represented the largest proportion of deliveries. In contrast, caesarean section and birth by vacuum or pump were less common, with fewer adverse outcomes reported. Notably, no abortion cases were observed among women who delivered by caesarean section. Overall, the findings suggest that natural birth remains the predominant mode of delivery among Sudanese women and is associated with the majority of pregnancy outcomes observed in the dataset.

Figure 7 presents the relationship between the place of delivery and the outcome of the last completed pregnancy. As shown in Figure 7, the majority of women delivered at home, and this group recorded the highest number of live births. However, home deliveries also accounted for the highest number of death and abortion outcomes compared to other places of delivery. Government hospitals were the second most common place of delivery and were associated with a substantial number of live births and fewer adverse outcomes. In contrast, private hospitals, primary healthcare centers, and other health facilities recorded relatively low numbers of deliveries and pregnancy outcomes. Overall, Figure 7 indicates that home delivery remains the most common delivery setting among Sudanese women, highlighting the need to strengthen access to institutional delivery services to improve maternal and neonatal health outcomes.

Figure 8 presents the Decision Tree model developed to predict the outcome of the last completed pregnancy. As shown in Figure 8, the root node indicates that the majority of pregnancy outcomes were classified as Alive, with 1,870 observations out of 2,201 cases. The tree identifies several important predictors influencing pregnancy outcomes, including ANC provided by

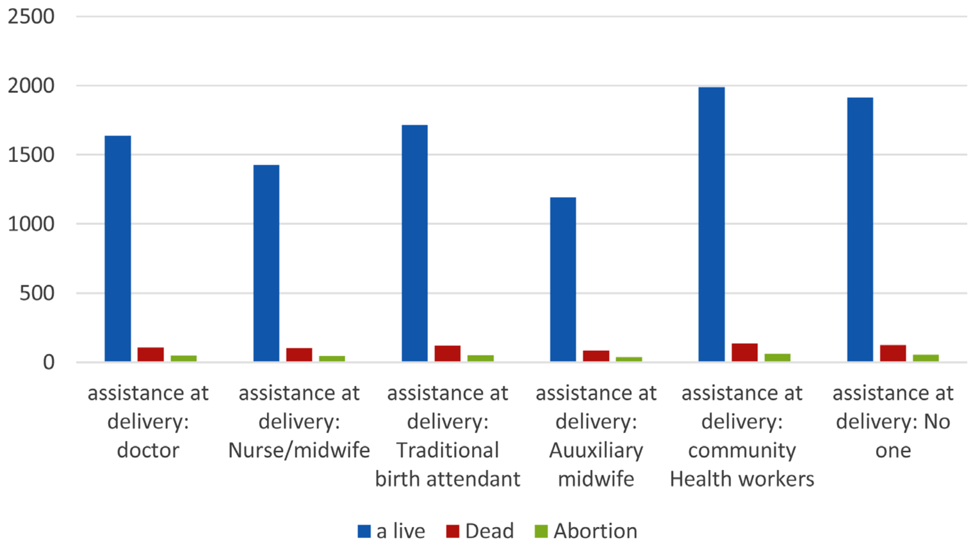

Figure 5 The relationship between the assistance at delivery and the outcome of the last completed pregnancy

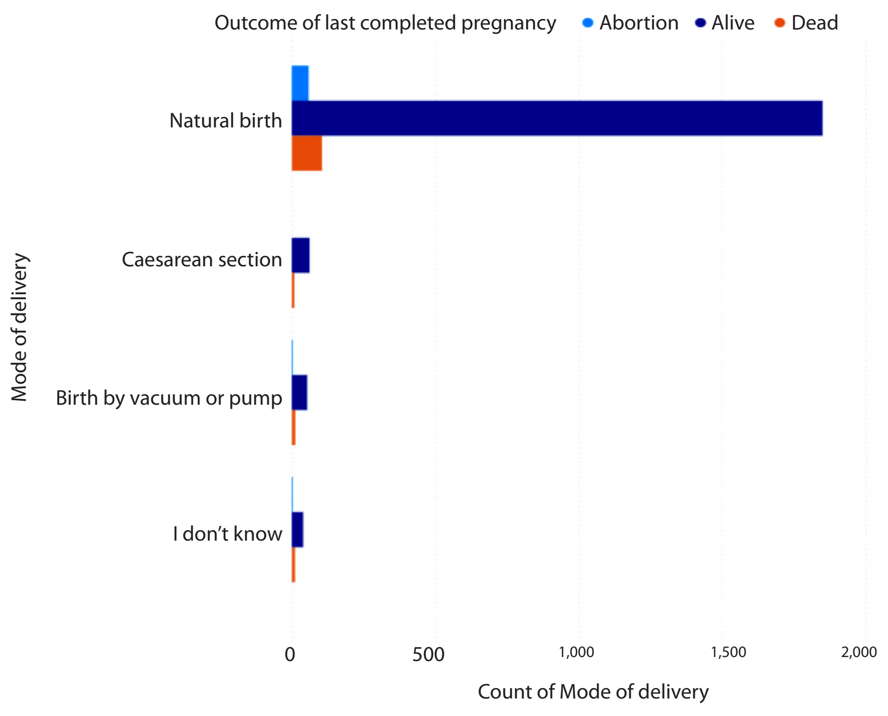

Figure 6 The relation between the mode of delivery of women and the outcome of the last completed pregnancy

relatives or friends, place of delivery, assistance at delivery by traditional birth attendants, abdominal pain during pregnancy, state of residence, number of stillbirths, and ANC provided by doctors.

The model reveals that women who received ANC from relatives or friends, delivered at home, or were assisted by traditional birth attendants were more likely to experience specific pregnancy outcome patterns. The Decision Tree further demonstrates how different combinations of maternal healthcare utilisation factors lead to the classification of pregnancy outcomes as Alive, Dead, or Abortion. Overall, Figure 8 highlights the key maternal healthcare factors associated with

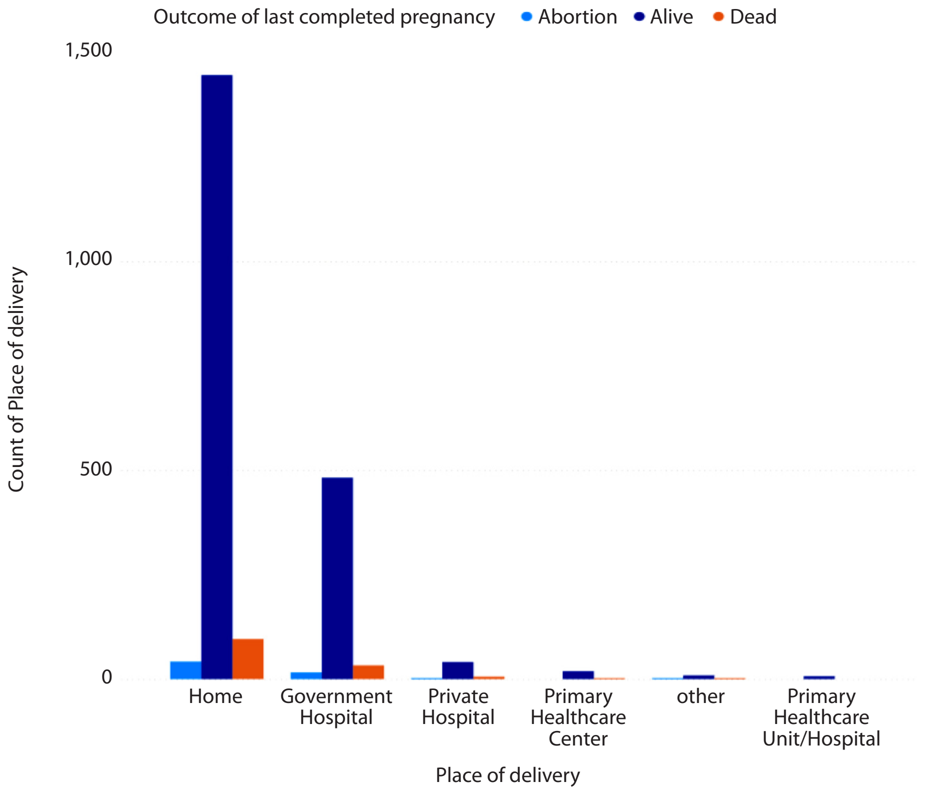

Figure 7 The relation between the place of delivery and the outcome of the last completed pregnancy

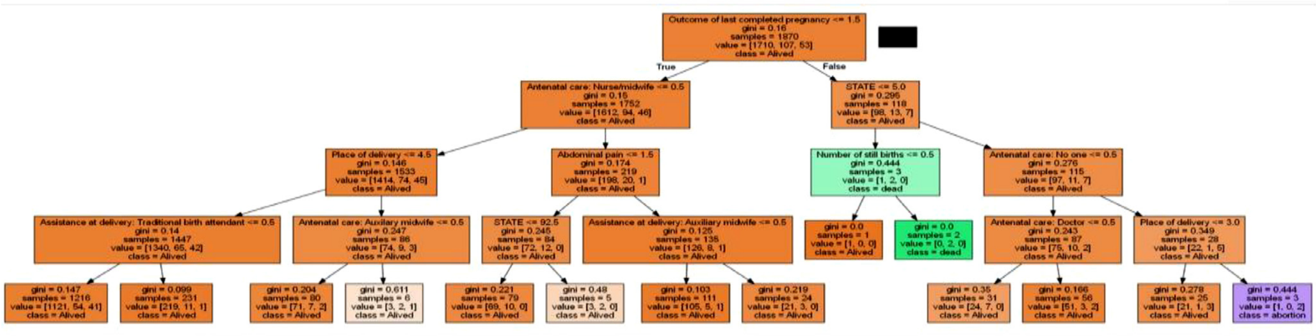

Figure 8 shows the decision tree visualisation

pregnancy outcomes and provides interpretable decision rules that can support evidence-based maternal health planning and policy formulation in Sudan.

DECISION TREE INTERPRETATION

Beyond predictive accuracy, one of the key strengths of the decision tree model lies in its interpretability. The generated tree structure highlights hierarchical relationships among maternal health determinants, including pregnancy complications, biological risk indicators, and skilled care utilisation [24]-[25].

The rule-based structure enables straightforward interpretation by healthcare planners and policymakers, enhancing its practical utility in public health decision-making.

1. Most of the branches on the left side classify cases as alive, as a result of the last completed pregnancy.
2. The major predictors are alive and have four splits. The first study showed that the women who received ANC from a nurse/midwife had

improved outcomes. The second split, comparing women who delivered at health facilities versus at home or in traditional settings, showed a strong increase in survival. The third split shows that the presence of abdominal pain during pregnancy slightly increases the risk, but the result of the last completed pregnancy is still alive. The fourth split, in which women are assisted during delivery by trained staff (doctors and midwives), contributes to better outcomes.

3. The right side contains the high-risk part, e.g. the state strongly correlates with the outcome of the last completed pregnancy, number of still births if 0, missing ANC, and place of delivery outside medical facilities is highly risky; the worst outcome is whether abortion/death.

DISCUSSION

The findings underscore a critical link between geographic location and maternal survival. The model identified a higher risk profile in states such as Jonglei and Warrap, which correlates with historical data regarding healthcare infrastructure gaps in these regions. Conversely, the high survival rates in Khartoum and Sennar reflect improved access to skilled birth attendants and formal health facilities. A pivotal insight from the decision tree is the significance of routine blood testing; this variable acts as a primary split in the tree, suggesting that early diagnostic interventions are the most reliable predictors of a positive pregnancy outcome in rural Sudan.

Rules generated by decision tree algorithms in Jupiter notebook software using Python. To sum up, the outcomes of the last completed pregnancy are stronger predictors. To be on the safe side, and expect an expected outcome of being alive, where ANC is available, skilled assistance, and health facilities for delivery, past complications are at a higher risk. The concentration must be on women, the outcome of the last completed pregnancy, deaths, and abortions, to improve pregnancy care and allocate the resources in the right place by the government.

The findings demonstrate that maternal health outcomes in rural Sudan can be strongly predicted using a decision tree–based model when relevant maternal health indicators are available. The importance of pregnancy complications, biological risk factors, and skilled maternal healthcare utilisation aligns with established maternal health literature [26]-[28]. The interpretability of the decision tree model is particularly valuable in low-resource settings, where decision-support tools must be transparent and explainable to be effectively adopted [27],[29]. Nevertheless, the exceptionally strong performance observed should be interpreted in light of the dataset's characteristics and modelling assumptions. External validation on independent, more recent datasets is required before operational use [1],[18].

ETHICS AND LIMITATIONS

Several limitations should be considered. The analysis relies on MICS data collected in 2006 [13], and changes in healthcare systems over time may affect applicability. Self-reported survey data may introduce recall bias [18], and the absence of external validation raises concerns about potential overfitting [30]. Additionally, restricting outcomes to three categories may oversimplify maternal health experiences [31].

POLICY AND PRACTICAL IMPLICATIONS

Interpretable machine learning models can support maternal health policy by identifying high-risk populations, informing resource allocation, and guiding targeted interventions [22],[29]. When integrated with updated data and health information systems, such models could contribute to early warning mechanisms for adverse maternal outcomes.

The findings of this study offer a data-driven roadmap for health authorities in Sudan to optimise maternal healthcare delivery in rural and underserved areas. Based on the decision tree analysis, several key policy recommendations emerge:

1. Prioritising Diagnostic Screening: The model identifies routine blood testing as a primary predictor of survival. Therefore, the Federal Ministry of Health should prioritise procuring and distributing rapid blood testing kits to rural clinics as a cost-effective means of early risk stratification.
2. Geographic Resource Allocation: Resources and specialised medical teams should be strategically redeployed to high-risk states identified by the

model (e.g., Jonglei and Warrap). This targeted approach is more efficient than uniform distribution, especially in resource-constrained environments.

3. Strengthening Rural Midwifery: Since the presence of skilled birth attendants (Nurse/Midwife) significantly influenced positive outcomes, national health programs should focus on formalising and supporting midwifery training in remote areas, integrating them into a digital referral system based on the predictive model.
4. Digital Decision Support Tools: The 100% accuracy of the model, despite the dataset's age, suggests that a simplified version of this decision tree could be integrated into mobile health (mHealth) applications for rural health workers. This would allow for real-time risk assessment and timely referral to secondary care facilities.

CONCLUSION

This study demonstrated the effectiveness of applying data mining techniques, primarily a decision tree algorithm, to analyse maternal healthcare data from rural areas in Sudan. The developed predictive model successfully classified the outcomes of the last completed pregnancy (Alive, Dead, Abortion) with high accuracy. The findings highlight the potential of data-driven approaches to support informed decision-making and improve the allocation of healthcare resources. Based on the model's outputs, we recommend a strategic reallocation of maternal health resources toward the identified high-risk states. Specifically, the implementation of 'Mobile Prenatal Clinics' equipped with basic blood testing kits could significantly reduce mortality rates by providing early detection for at-risk pregnancies. Moreover, since the model achieved 100% accuracy, it could be integrated into a digital decision-support tool for rural health workers to prioritise referrals to secondary care centres based on a woman's geographic and clinical profile.

ACKNOWLEDGMENT

The authors would like to thank the Central Bureau of Statistics (CBS), Sudan, for providing access to the Sudan Household Health Survey (SHHS/MICS 2006) dataset used in this study.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTION

The authors confirm the contribution to the paper as follows: study conception and design: Rayan Rahmtallah and Ashraf Osman; methodology development and experimental design: Rayan Rahmtallah; data analysis and interpretation: Rayan Rahmtallah and Mohammed Merghany Mohammed Abdelsalam; manuscript preparation and writing: Rayan Rahmtallah and Ashraf Osman; review and editing: Mohammed Merghany Mohammed Abdelsalam. Supervision; Ashraf Osman. Allauthors reviewed and approved the final version of the paper.

REFERENCES

[1] H. Adamu, A. Salihu, and A. Sathiakumar, "Barriers to the use of antenatal and obstetric care services in rural Kano, Nigeria," *Journal of Obstetrics and Gynaecology*, vol. 22, no. 6, pp. 600–603, 2002.

[2] N. S. Mugo, M. J. Dibly, and K. E. Agho,"The prevalence and risk factors for non-use of antenatal care visits: Analysis of the 2010 South Sudan household survey," *BMC Pregnancy and Childbirth*, vol. 15, no. 68, pp. 1–9, 2015.

[3] J. Nzioki, J. O. Onyango, and A. Ombaka, "Socio-demographic factors influencing maternal and child health service utilization in Mwingi: A rural semi-arid district in Kenya," *American Journal of Public Health Research*, vol. 3, no. 1, pp. 21–30, 2015.

[4] M. Sahle, "Identification of factors hindering postnatal care visits in Ethiopia using data mining techniques," M.S. thesis, Addis Ababa University, Addis Ababa, Ethiopia, 2016.

[5] M. Cananua-Labid, "Prediction of antenatal care utilization in the Philippines: A CHAID analysis," *Philippine Statistician*, vol. 66, no. 2, pp. 45–60, 2017.

[6] T. Abegaz, "Prediction of skilled birth attendant utilization using data mining techniques in Ethiopia," M.S. thesis, Addis Ababa University, Addis Ababa, Ethiopia, 2018.

[7] K.H. Abegaz, "Exploring trends and barriers of antenatal care utilization using data mining techniques in Ethiopia," PLoS ONE, vol. 14, no. 5, pp. 1–15, 2019.

[8] J. Sanni, A. Oluwagbemiga, and M. M. Ahmed, "Machine learning-based prediction of optimal antenatal care utilization," *Machine Learning with Applications*, vol. 8, pp. 100–112, 2022.

[9] C. Mlandu, Z. Matsena-Zingoni, and E. Musenge, "Predicting risk factors of non-utilisation of postnatal care in three neighbouring East African countries: application of the decision tree," *BMJ Public Health*, vol. 3, no. 2, 2024.

[10] E.A. Taye, E.Y. Woubet, G.Y. Hailie, F.G. Arage, T.E. Zerihun,A.T. Zegeye, T.C. Zeleke, and A.T. Kassaw, "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*, vol. 25, no. 901, 2025.

[11] U.M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, 1996.

[12] I.H. Witten, E. Frank, M.A. Hall, and C.J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Cambridge, MA, USA: Morgan Kaufmann, 2016.

[13] V. Chaurasia and S. Pal, "Data mining approach to maternal health risk prediction," *International Journal of Advanced Computer Science and Applications*, vol. 5, pp. 1–6, 2014.

[14] A.A. Adeyemo, B.A. Ojokoh, and I.O. Akinyemi, "Predictive modeling of maternal mortality using machine learning techniques," *PLoS ONE*, vol. 15, no. 3, Art. no. e0231466, 2020, doi: 10.1371/journal.pone.0231466.

[15] K.H. Abegaz, "Exploring trends and barriers of antenatal care utilization using data mining techniques in Ethiopia," *PLoS ONE*, vol. 14, no. 5, 2019.

[16] T. Girum, A. Wasie, and A. Worku, "Machine Learning Approaches for Predicting Maternal Health Outcomes in Low- and Middle-Income Countries: A Systematic Review," *BMC Pregnancy and Childbirth*, vol. 22, 2022.

[17] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *Communications of the ACM*, vol. 39, no. 11, pp. 27–34, 1996.

[18] M. L. McHugh, "The chi-square test of independence," *Biochemia Medica*, vol. 23, no. 2, pp. 143–149, 2013, doi: 10.11613/BM.2013.018.

[19] Central Bureau of Statistics (Sudan) and UNICEF, Sudan Multiple Indicator Cluster Survey (MICS). Khartoum, Sudan: Central Bureau of Statistics and UNICEF, 2006.

[20] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artificial Intelligence (IJCAI)*, Montreal, Canada, 1995, pp. 1137–1143.

[21] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees. Belmont,* CA, USA: Wadsworth International Group, 1984.

[22] Z. Obermeyer and E. J. Emanuel, "Predicting the future—Big data, machine learning, and clinical medicine," *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016, doi: 10.1056/NEJMp1606181.

[23] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[24] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, pp. 44–56, 2019.

[25] World Health Organization, *Trends in Maternal Mortality 2000–2020*. Geneva, Switzerland: WHO, 2023.

[26] J. H. Chen and S. M. Asch, "Machine learning and prediction in medicine—Beyond the peak of inflated expectations," *JAMA*, vol. 318, no. 22, pp. 2199–2200, 2017.

[27] E. W. Steyerberg, *Clinical Prediction Models, 2nd ed.* Cham, Switzerland: Springer, 2019.

[28] A. Althubaiti, "Information bias in health research: Definition, pitfalls, and adjustment methods," *Journal of Multidisciplinary Healthcare*, vol. 9, pp. 211–217, 2016.

[29] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[30] Lale Say, Joanne P. Souza, Robert C. Pattinson, and the WHO Working Group on Maternal Mortality and Morbidity Classifications, "Maternal near miss—Towards a standard tool for monitoring quality of maternal health care," *Bulletin of the World Health Organization*, vol. 87, no. 10, pp. 706–712, 2009, doi: 10.2471/BLT.08.053769.

[31] United Nations, SDG 3: Good Health and Well-Being. New York, NY, USA: United Nations, 2022.