Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Optimizing machine learning models for predicting iron supplementation uptake among pregnant women in Somaliland: insights from the 2020 Somaliland demographic and health survey data

Abdifatah Ibrahim Mouse

abdifatahibrahim@amoud.edu.so

Amoud University

Omran Salih

Durban University of Technology

Abdisalam Hassan Muse

Amoud University

Article

Keywords: Iron Supplementation, Pregnant Women, Machine Learning, Random Forest, Somaliland

Posted Date: March 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8431473/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Optimizing machine learning models for predicting iron supplementation uptake among pregnant women in Somaliland: insights from the 2020 Somaliland demographic and health survey data

Abdifatah Ibrahim Mouse[1]*, Omran Salih[2], and Abdisalam Hassan Muse[1,2]
Email: abdifatahibrahim@amoud.edu.so, Email: omran.salih@amoud.edu.so and Email: abdisalam.hassan@amoud.edu.so

***Correspondence:** Abdifatah Ibrahim Mouse
Email: abdifatahibrahim@amoud.edu.so

Abstract

Iron supplementation during pregnancy is crucial, fulfilling increased demands for placental and fetal development. Despite WHO recommendations and efforts to promote iron intake, uptake remains suboptimal in many regions, including Somaliland, where maternal and child health indicators are poor due to limited healthcare access and nutritional deficiencies. This study aims to identify determinants of iron supplementation to inform targeted interventions. This cross-sectional study utilized data from the 2020 Somaliland Health and Demographic Survey (SLHDS) for a sample of 2,983 pregnant women. Explanatory variables included maternal age, education, employment status, ANC visits, residence type, region, media exposure, and wealth. Supervised machine learning models including, Logistic Regression, Random Forest, XGBoost, LightGBM, Support Vector Machine, and K-Nearest Neighbors were employed to predict iron supplementation uptake. Performance was evaluated using accuracy, precision, recall, F1-score, and AUROC. Overall, 28.83% of pregnant women reported taking iron supplements. Bivariate analysis revealed significant associations ($p < 0.05$) between iron supplementation and maternal age ($\chi2 = 15.00$, $p = 0.020$), educational level ($\chi2 = 117.3$, $p < 0.001$), employment status ($\chi2 = 5.5$, $p = 0.019$), ANC visits ($\chi2 = 259.5$, $p < 0.001$), region ($\chi2 = 103.5$, $p < 0.001$), media exposure ($\chi2 = 22.3$, $p < 0.001$), and wealth quintile ($\chi2 = 261.1$, $p < 0.001$). The Random Forest model demonstrated the best performance, achieving an accuracy of 0.785 and an AUROC of 0.81. Iron supplementation uptake in Somaliland remains suboptimal, with only 28.83% reporting adherence, underscoring a critical need for enhanced interventions. The Random Forest model highlighted key predictors of iron supplementation uptake: wealth status, region, antenatal care visits, and maternal age. These findings emphasize the importance of socioeconomic factors, geographical location, and access to healthcare services in influencing iron supplementation behaviors.

Keywords: Iron Supplementation, Pregnant Women, Machine Learning, Random Forest, Somaliland

1. Introduction

Iron is an essential mineral that supports oxygen transport in the blood, plays a key role in various bodily functions, and is commonly used to prevent and treat iron-deficiency anemia. Anemia, which is frequently caused by iron deficiency, is a serious global public health issue, especially for pregnant women in developing nations (1,2). In order to improve the health of both the mother and the fetus, iron supplementation—the administration of extra iron—is a critical strategy in defending against iron deficiency and anemia (3–5). The primary goal of iron supplementation during pregnancy is to fulfill the higher iron needs required for the growth of the placenta, fetal development, and the expansion of the mother's red blood cell mass (4,6). Supplementing to address iron deficiency is essential since low iron levels during pregnancy can have serious repercussions, such as an increased risk of maternal death, obstetric problems, preterm birth, and low birth weight babies (7–9). Regardless of hemoglobin levels, the World Health Organization (WHO) recommends iron supplementation for all pregnant women due to the potential of such negative health effects (10).

Globally, anemia is a widespread health issue among pregnant women, with an estimated global prevalence of 41.8% (1). A significant portion of this burden is attributed to iron deficiency (2,11). In 2019, the WHO reported that 29.9% of women of reproductive age worldwide were anemic, with a higher prevalence among pregnant women (36.5%) compared to their non-pregnant counterparts (29.6%) (12). Sub-Saharan Africa (SSA) bears a disproportionately high burden of anemia during pregnancy, with prevalence rates exceeding those in other

regions (13). While adherence to iron supplementation in high-income countries can reach up to 85%, rates in some SSA countries have been reported to be as low as 22% (12,14). In Ethiopia, for example, a spatial analysis of iron supplementation among pregnant women revealed significant regional variations in coverage (15). Lower-middle-income countries, observed the highest prevalence of anemia.

In the East African region, anemia among pregnant women remains a persistent challenge. The coverage of iron supplementation during pregnancy is still low, and has not fulfilled the World Health Organization (WHO) standard recommendations (16). Somaliland is a self-declared independent state located in the Horn of Africa. Maternal and child health indicators are particularly poor in Somaliland, reflecting limited access to healthcare services and pervasive nutritional deficiencies (17). Somaliland is facing with poor maternal and child health indicators compared to other countries in the region (18). Although routine iron supplementation is recommended as part of antenatal care, uptake remains suboptimal. Understanding the factors influencing iron supplementation among pregnant women in Somaliland is therefore crucial for designing targeted interventions to improve maternal and child health outcomes. Somaliland's healthcare system faces constraints such as limited resources, inadequate infrastructure, and a shortage of skilled healthcare workers (19). Further, cultural beliefs, socioeconomic disparities, and geographical barriers may also affect access to and utilization of maternal health services.

Despite efforts to promote iron supplementation during pregnancy, the specific factors influencing its uptake in Somaliland remain poorly understood. Existing literature from other regions, including Ethiopia, and Tanzania, highlights the importance of factors such as antenatal care attendance, maternal education, household wealth, media exposure, and community-level factors (15,20). However, the unique context of Somaliland, with its distinct cultural norms, healthcare system challenges, and socioeconomic conditions, necessitates a focused investigation to identify the key determinants of iron supplementation in this specific setting.

This study aims to assess the prevalence of iron supplementation among pregnant women in Somaliland and to predict individual and community-level factors associated with its uptake, utilizing data from the 2020 Somaliland Health and Demographic Survey (SLHDS 2020). By looking at these variables, the study aims to offer evidence-based suggestions for enhancing iron supplementation programs, which would ultimately help to lower maternal anemia and improve the health of mothers and children in Somaliland.

2. Methods and Materials

2.1 Study Area

Somaliland, situated in the Horn of Africa, shares borders with Djibouti to the west, Ethiopia to the south, and Somalia to the east. Encompassing 176,119.2 square kilometers, the country possesses an 850-kilometer coastline along the Gulf of Aden. The climate varies regionally, with the northern terrain characterized by hills and altitudes ranging from 1,800 to 2,100 meters above sea level. While Awdal, Marodijeh, and portions of Sahil exhibit fertile and mountainous landscapes, Togdheer, Sanaag, and Sool are predominantly semi-arid, experiencing average daily temperatures between 25°C and 35°C and humidity fluctuations from 63% during the dry season to 82% during the wet season. The country experiences four distinct seasons: Gu' (spring), Hagaa (summer), Dayr (autumn), and Jiilaal (winter).

2.2 Data source, Study Design, Period and Sampling procedure

In this study, we used the most recent Somaliland Health and Demographic Surveys (SLHDSs) 2020 dataset as the source of the data for this study. The SLHDS2020 is the first countrywide household survey that collected demographic and health data in all six main regions of Somaliland.

The SLHDS 2020 employed a cross-sectional study design. Cross-sectional studies provide a snapshot of the population at a single point in time, allowing for the examination of associations between the study variables (21). Data collection was conducted from November 22, 2019, to March 1, 2020. A stratified, multi-stage probability cluster sampling design was applied across six geographical regions, each stratified by urban, rural, and nomadic residence areas. The sample design utilized up-to-date high-resolution satellite imagery to digitize all dwelling structures in urban and rural areas, forming Enumeration Areas (EAs) on-screen using GIS software. Each EA contained a minimum of 50 and a maximum of 149 dwelling structures. A total of 2,923 EAs (1,869 in urban areas and 1,054 in rural areas) were digitized. In the first stage, 35 EAs in every stratum of every design domain were selected using probability proportional to size (PPS) sampling of digitized dwelling structures.

Listing of households was carried out in each of the 35 selected EAs to obtain the total number of households. In the second stage, 10 PSUs were sampled using probability proportional to the number of listed households.

2.3 Population and Sample

The target population for this study comprises women of reproductive age (15-49) who were pregnant at the time of the SLDHS 2020 survey. The SLHDS successfully total Number of eligible women interviewed 6,285 in this age group. From this sample, we extracted data for all women who reported being currently pregnant. After excluding cases with missing data on the outcome variable (iron supplement use) and other key predictors, a final analytical sample of 2,983 pregnant women was included in the machine learning analysis.

2.4 Inclusion and Exclusion Criteria

The inclusion criterion for this study was any woman aged 15-49 who was currently pregnant at the time of the survey and had complete information on iron supplement use. The primary exclusion criteria were: (i) women who were not pregnant at the time of the survey, and (ii) pregnant women with missing values for the outcome variable or the key explanatory variables selected for the models, to ensure the robustness of the analysis.

2.5 Target Variable

The primary response variable in this study was uptake of iron supplementation during pregnancy. This was measured through a direct question in the SLHDS 2020 questionnaire, asking women whether they had taken iron supplements during their current or most recent pregnancy. The response was coded as a binary variable (Yes=1, No=0), indicating whether or not the woman had used iron supplements.

2.6 Explanatory Variables

Based on existing literature on determinants of iron supplement use, and guided by the availability of variables in the SLDHS 2020, we included the following explanatory variables in the analysis: maternal age (15–19, 20–24, 25–29, 30–34, 35–39, 40–44, and 45–49 years), maternal education (no education, primary, secondary, and higher), employment status (yes or no), number of ANC visits (0–4 visits vs. >4 visits), type of residence (rural or urban), region (Awdal, Marodijeh, Sahil, Togdheer, Sool, and Sanaag), media exposure (no exposure vs. exposure), and wealth status (poor, middle, rich).

2.7 Operational Definitions

a) **Iron Supplement Use:** Consumption of iron supplements (tablets or syrup) at any point during the current or most recent pregnancy (22).
b) **Maternal Education:** The highest level of formal education completed by the pregnant woman (No Education, Primary, Secondary, Higher) (23).
c) **Antenatal Care (ANC) Visits:** The number of times a pregnant woman received health care from a skilled provider (doctor, nurse, or midwife) during her current pregnancy. For this study, it was categorized as 'None', '1-3 visits', and '4 or more visits', in line with WHO recommendations (24).
d) **Wealth Status:** A composite measure of household economic status, derived from household assets and characteristics using principal component analysis (quintiles: poorest, poorer, middle, richer, and richest) (25).
e) **Media Exposure:** Whether the pregnant woman reported being exposed to any form of media (television, and radio).

2.8 Data Processing

Data processing was performed using STATA (version 16.0) and R (version 4.2). The raw SLHDS dataset was cleaned by selecting the relevant variables and managing missing values through deletion due to the small proportion of missing data on key predictors. The dataset exhibited a class imbalance in the outcome variable, which can bias machine learning models. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training dataset to create a balanced distribution between the iron supplemented and 'Not supplemented' classes (26). The entire dataset was split into a training set (80%) and a

testing set (20%), using stratified sampling to maintain the original distribution of the outcome variable in both subsets.

2.9 Necessary Libraries

The following R libraries were utilized for data analysis and machine learning modeling: haven, caret, randomForest, xgboost, glmnet, e1071, nnet, class, gbm, rpart, lightgbm, pROC, themis, recipes, dplyr, Matrix, ggplot2, reshape2, mlbench, SHAPforxgboost, and shapviz. These libraries provide a wide range of functions for data manipulation, statistical analysis, machine learning, and visualization.

2.10 Statistical Analysis

The analysis was conducted in three stages: univariate, bivariate, and machine learning modeling.

2.10.1 Univariate Analysis

Descriptive statistics, including frequencies and percentages, were calculated to summarize the socio-demographic and health-related characteristics of the study participants.

2.10.2 Bivariate Analysis

Pearson's chi-square (χ^2) test was used to assess the association between each categorical explanatory variable and the outcome variable (iron supplement use). A p-value of <0.05 was considered statistically significant. The formula for the chi-square test is:

$$\chi^2 = \left[\sum \frac{(O_i - E_i)^2}{E_i}\right]$$

where O_i is the observed frequency and E_i is the expected frequency for each cell in the contingency table.

2.10.3 Machine Learning Analysis

Several supervised machine learning classification algorithms were employed to identify the most important predictors of iron supplement use. The models included Logistic Regression (LR), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) (27–30).

Model performance was evaluated on the test set using standard metrics derived from the confusion matrix: Accuracy, Precision, Recall (Sensitivity), F1-Score, Matthews Correlation Coefficient (MCC), and the Area Under the Receiver Operating Characteristic Curve (AUROC) (31).

- **Accuracy** = $\frac{(TP + TN)}{(TP + TN + FP + FN)}$
- **Precision** = $\frac{TP}{(TP + FP)}$
- **Recall (Sensitivity)** = $\frac{TP}{(TP + FN)}$
- **F1-Score** = $\frac{2 * (Precision * Recall)}{(Precision + Recall)}$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

2.11 Ethical approval

This study utilized publicly accessible and fully anonymized data from the Somaliland Health and Demographic Survey (SLHDS) 2020, which contains no personally identifying information.

3. Results

3.1 Prevalence of Iron Supplementation Uptake

A total of 2,983 pregnant women were included in this analysis from the 2020 SLDHS data. Overall, 28.83% (n = 860) of pregnant women reported taking iron supplements (Fig. 1). Among those who had 0-4 antenatal care (ANC) visits (91.59%), are majority of the respondents, while only 8.4% visits more than times (Fig. 2).

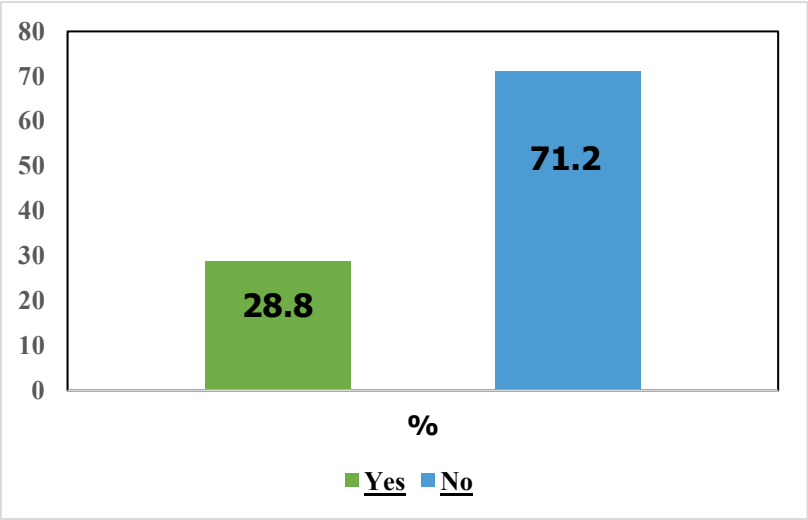


Figure 1. Prevalence of Iron Supplementation Among Pregnant Women (N = 2,983)

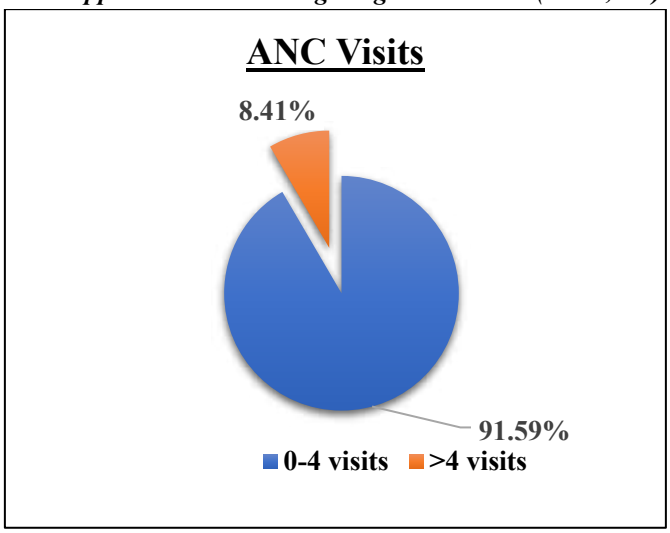


Figure 2. Number of Antenatal Visit Care Among Pregnant Women (N = 2,983)

3.2 Descriptive outcomes of the background characteristics

The study sample (N = 2,983) of pregnant women in Somaliland was predominantly composed of individuals aged 25-29 years (26.79%), with no formal education (83.04%), who were unemployed (99.53%). The majority of participants had no media exposure (98.12%), and were classified as poor (54.68%).

Bivariate analyses, revealed statistically significant associations between iron supplementation and several individual and community-level factors:

Women aged 15–19 years reported the highest uptake of iron supplements (31.55%), and maternal age was significantly associated with supplementation, $\chi^2(6) = 15.00$, $p = .020$. Educational attainment also demonstrated a strong association, $\chi^2(3) = 117.29$, $p < .001$, with the highest uptake observed among women with higher education (66.67%). Employment status was significantly related to iron supplementation, $\chi^2(1) = 5.50$, $p = .019$, as employed women were more likely to take iron supplements (57.14%) compared to unemployed women.

Antenatal care (ANC) visits showed a highly significant association, $\chi^2(1) = 259.51$, $p < .001$. Women who attended more than four ANC visits had substantially higher supplementation rates (72.91%) than those with fewer visits. Regional differences were also statistically significant, $\chi^2(5) = 103.49$, $p < .001$, with the highest uptake found in Sahil (38.36%), Marodijeh (36.72%), and Togdheer (36.58%). Media exposure was positively associated with iron supplementation, $\chi^2(1) = 22.30$, $p < .001$; women with media exposure had a higher uptake rate (57.14%) compared to those with no exposure (28.29%). Similarly, wealth status was significantly related to iron supplementation, $\chi^2(2) = 261.10$, $p < .001$, with the richest group showing the highest uptake (46.27%).

Table 1. Descriptive Statistics: Frequencies, Percentages and Chi-square Test Results with p-values.

Variable	Total	Uptake of Iron Supplementation		χ^2	*p*
	n (%)	No (N = 2,123)	Yes (860)		
Age Group					
15-19	168 (5.63%)	115 (68.45%)	53 (31.55%)		
20-24	549 (18.40%)	400 (72.86%)	149 (27.14%)		
25-29	799 (26.79%)	552 (69.09%)	247 (30.91%)		
30-34	632 (21.19%)	432 (68.35%)	200 (31.65%)	15.00	0.02
35-39	528 (17.70%)	382 (72.35%)	146 (27.65%)		
40-44	231 (7.74%)	180 (77.92%)	51 (22.08%)		
45-49	76 (2.55%)	62 (81.58%)	14 (18.42%)		
Maternal Education					
No Education	2,477 (83.04%)	1,860 (75.09%)	617 (24.91%)		
Primary	398 (13.34%)	217 (54.52%)	181 (45.48%)	117.3	<.001
Secondary	75 (2.51%)	35 (46.67%)	40 (53.33%)		
Higher	33 (1.11%)	11 (33.33%)	22 (66.67%)		
Employment Status					
Yes	14 (0.47%)	6 (42.86%)	8 (57.14%)	5.5	0.02
No	2,969 (99.53%)	2,117 (71.30%)	852 (28.70%)		
ANC Visits					
0-4 visits	2,732 (91.59%)	2,055 (75.22%)	677 (24.78%)	259.5	<.001
>4 visits	251 (8.41%)	68 (27.09%)	183 (72.91%)		
Media Exposure					
No Exposure	2,927 (98.12%)	2,099 (71.71%)	828 (28.29%)	22.3	<.001
Exposure	56 (1.88%)	24 (42.86%)	32 (57.14%)		
Wealth Status					
Poor	1,631 (54.68%)	1,350 (82.77%)	281 (17.23%)	261.1	<.001
Middle	321 (10.76%)	219 (68.22%)	102 (31.78%)		
Rich	1,031 (34.56%)	554 (53.73%)	477 (46.27%)		
Region					
Awdal	376 (12.60%)	247 (65.69%)	129 (34.31%)		
Marodijeh	354 (11.87%)	224 (63.28%)	130 (36.72%)		
Sahil	365 (12.24%)	225 (61.64%)	140 (38.36%)	103.5	<.001
Togdheer	514 (17.23%)	326 (63.42%)	188 (36.58%)		
Sool	678 (22.73%)	556 (82.01%)	122 (17.99%)		
Sanaag	696 (23.33%)	545 (78.30%)	151 (21.70%)		
Type of Residence					
Rural	1,489 (49.92%)	1,065 (71.52%)	424 (28.48%)	0.18	0.67
Urban	1,494 (50.08%)	1,058 (70.82%)	436 (29.18%)		

3.3 Model performance comparison for predicting iron supplementation uptake

The predictive capabilities of ten machine learning models, demonstrating a range of performance, were evaluated for identifying predictors of iron supplement use among pregnant women, with performance metrics summarized in Table 2. The Random Forest (RF) model achieved the highest accuracy (0.785, 95% CI [0.750, 0.818]) and AUC (0.81), suggesting a superior ability to correctly classify individuals and discriminate between users and non-users of iron supplements. Logistic Regression (LR) achieved an accuracy of 0.705 (95% CI [0.666, 0.741]) and an AUC of 0.756, representing moderate predictive power. Support Vector Machine (SVM) had the lowest accuracy (0.290, 95% CI [0.254, 0.329]) and relatively low sensitivity (0.264), indicating limited effectiveness in identifying iron supplement users. K-Nearest Neighbors (k-NN) attained an accuracy of 0.763 (95% CI [0.727, 0.797]), and a Kappa coefficient of 0.427, suggesting moderate agreement between predicted and actual classifications. Decision Tree (DT) model achieved an accuracy of 0.722 (95% CI [0.684, 0.757]) and an AUC of 0.770, demonstrating reasonable, but not exceptional, predictive performance. Gradient Boosting

Machine (GBM) and XGBoost model achieved an accuracy of 0.757 (95% CI [0.720, 0.791]) and 0.770 (95% CI [0.734, 0.803]) respectively. Neural Network achieved an accuracy of 0.727 (95% CI [0.689, 0.762]) and LightGBM achieved an accuracy of 0.765 (95% CI [0.729, 0.799]). Naive Bayes exhibited a high sensitivity (0.920) but low specificity (0.384), indicating a tendency to over-predict iron supplement use, with an accuracy of 0.765 (95% CI [0.729, 0.799]). Considering the balance between sensitivity and specificity, as well as overall accuracy and discriminatory ability (AUC), the Random Forest model emerged as the most promising for identifying predictors of iron supplement use in this population.

The receiver operating characteristic (ROC) curves (Figure 2) illustrate that Random Forest (RF) model demonstrated the highest area under the curve (AUC = 0.81), indicating superior discriminatory ability compared to other models, followed by Logistic Regression (AUC = 0.76). Support Vector Machine (AUC = 0.74) and Naive Bayes (AUC = 0.74) had similar performance. Decision Tree, Gradient Boosting Machine, LightGBM, and XGBoost had similar AUC (0.77). Neural Network and KNN has less AUC compared to other models.

Table 2. Accuracy, confidence interval, kappa, sensitivity, specificity, precision, F1-measure, and AUC for the machine learning algorithms

Metric	**Random Forest**	**Logistic Regression**	**SVM**	**k-NN**	**Decision Tree**
Accuracy	0.785	0.705	0.290	0.763	0.722
95% CI	(0.750, 0.818)	(0.666, 0.741)	(0.254, 0.329)	(0.727, 0.797)	(0.684, 0.757)
Kappa	0.464	0.360	-0.283	0.427	0.368
Sensitivity	0.866	0.708	0.264	0.830	0.755
Specificity	0.587	0.698	0.355	0.599	0.640
Precision (PPV)	0.838	0.852	0.502	0.836	0.838
F1 Score	0.852	0.773	0.346	0.833	0.794
AUC	0.808	0.756	0.745	0.715	0.770
Metric	**GBM**	**XGBoost**	**Neural Network**	**LightGBM**	**Naive Bayes**
Accuracy	0.757	0.770	0.727	0.765	0.765
95% CI	(0.720, 0.791)	(0.734, 0.803)	(0.689, 0.762)	(0.729, 0.799)	(0.729, 0.799)
Kappa	0.411	0.425	0.384	0.414	0.347
Sensitivity	0.826	0.856	0.752	0.851	0.920
Specificity	0.587	0.558	0.663	0.552	0.384
Precision (PPV)	0.831	0.827	0.846	0.824	0.786
F1 Score	0.828	0.841	0.797	0.838	0.848
AUC	0.774	0.774	0.755	0.772	0.744

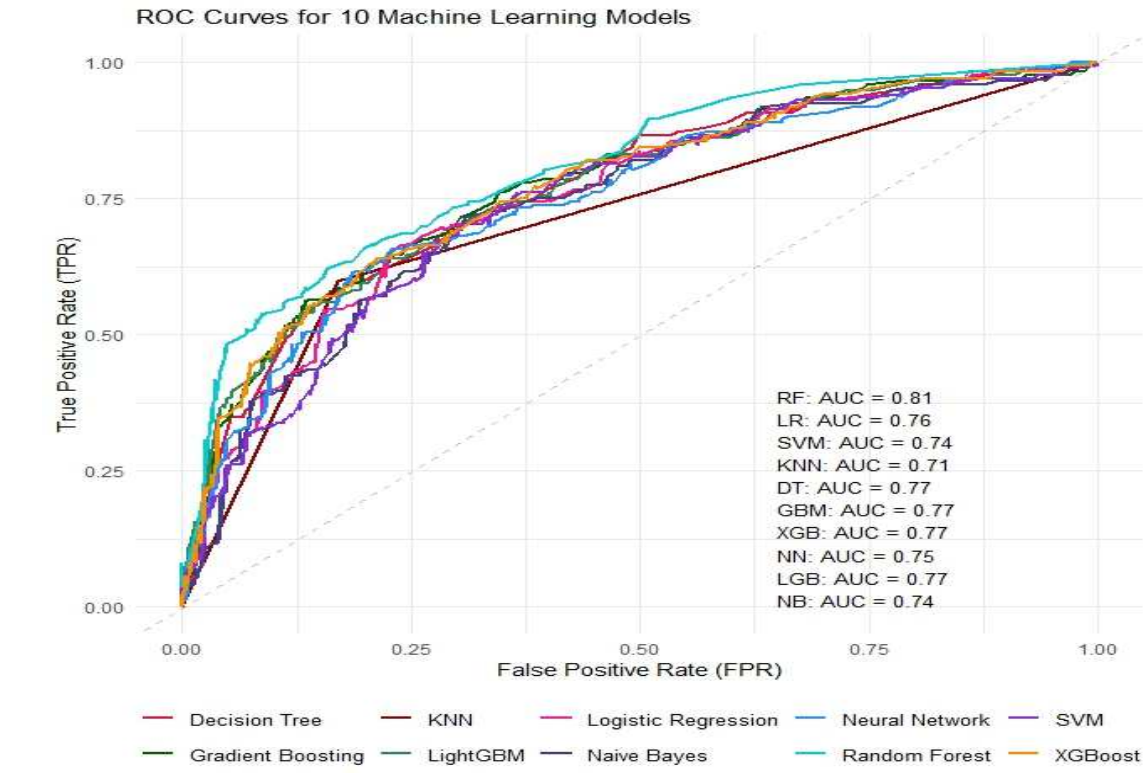


Figure 3. AUC curve for machine learning models

3.4 Features importance

Figure 3 presents the feature importance rankings derived from the Random Forest model. Wealth status emerged as the most influential predictor of iron supplement use, followed by region and the number of antenatal care visits. Maternal education, age group, and place of residence demonstrated moderate predictive importance. In contrast, media access and employment status exhibited relatively low importance, indicating a lesser contribution to the prediction of iron supplement utilization.

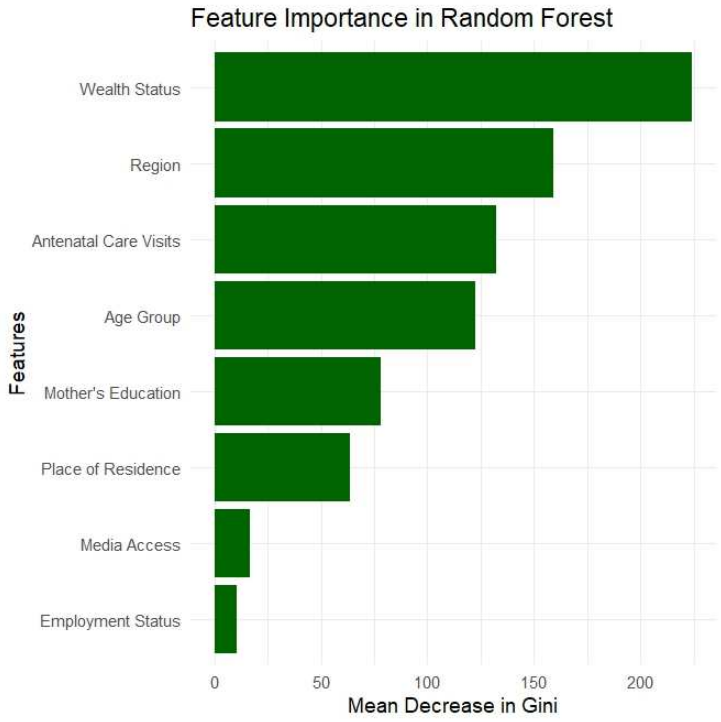


Figure 4. Important features selected from the dataset

4. Discussion

In this study, the main purpose was to (i) determine the individual and community-level factors associated with the uptake of iron supplementation in Somaliland and (ii) predict the iron supplementation and identify their key factors among pregnant women using a supervised machine learning model.

The prevalence of iron supplement uptake in this study was 28.8%. This finding aligns with a previous study done in 22 Sub-Saharan African countries (13). The prevalence of iron supplementation uptake observed in this study contrasts with several studies conducted across different geographical contexts. While Boti et al. (32) reported a substantially higher prevalence of 81.6% among pastoralist women in Southern Ethiopia, and Zegeye et al. (20) found a prevalence of 51.7% across 25 Sub-Saharan African countries, the current results are notably higher than rates documented in Pakistan: 38.3% (33) and 30% (34). Conversely, the present prevalence is lower than the 88.84% reported in Indonesia by Efendi et al. (12), the 77% found in Denmark by Knudsen et al. (35), and the 70% observed in the Philippines by Lutsey et al. (36). These disparities highlight significant geographical and contextual variations in adherence patterns, likely influenced by differences in health policies, antenatal care coverage, cultural practices, and health education efforts. The prevalence of iron supplementation uptake among pregnant women in Somaliland is significantly higher than that reported in several neighboring East African countries. For example, Kiwanuka et al. (37) documented an adherence rate of 12% at Mulago National Referral Hospital in Uganda. In Ethiopia, Haile et al. (38) reported a national adherence rate of 17.1%, while Ogundipe et al. (14) observed a rate of 22.3% among hospital-based participants in Northern Tanzania. These regional variations in supplementation practices may be attributed to differences in supplement availability, national health policies, antenatal care programs, the involvement of community health workers, culturally specific health beliefs, and public health awareness campaigns all of which may have positively influenced maternal nutrition practices in Somaliland.

The use of supervised machine learning models, such as Random Forest, Decision Tree, Logistic Regression, Naive Bayes, Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Gradient Boosting Machine (GBM), XGBoost, LightGBM, and Neural Network, enabled the development of a robust and accurate predictive model. For this objective, ten supervised machine-learning algorithms were trained on balanced training data using a train-test split. Accuracy, 95% CI, Kappa, Sensitivity, Specificity, Precision (PPV), F1 Score, AUC were used to compare the performance of classifier models. The Random Forest model outperformed other machine learning algorithms in predicting the outcome, achieving the highest accuracy of 78.5%, an area under the curve (AUC) of 80.8%, sensitivity of 86.6%, and an F1 score of 85.2%. These results indicate that Random Forest provided the most robust and reliable predictions compared to the other models evaluated.

These findings align with previous studies that have employed machine learning models to predict health-related outcomes. For instance, a study conducted in East African countries reported a high performance of the Random Forest classifier with an AUC of 89.2% and an accuracy of 94% (39). Similarly, another study across 27 Sub-Saharan African countries achieved an accuracy of 92%, precision of 93%, recall of 96%, and an F1 score of 94% (40). In Senegal, an AUC of 80.45%, F1 score of 75.89%, and an accuracy of 73.39% (41). Furthermore, a study in Bangladesh demonstrated outstanding results, with an accuracy of 98.75%, recall of 98.61%, precision of 98.88%, F1 score of 98.74%, and an AUROC of 99.79% (42). These studies support growing evidence that machine learning algorithms particularly ensemble models like Random Forest are effective in predicting health behaviors and outcomes across various settings.

In the present study, Random Forest model identified wealth status, region, antenatal care (ANC) visits, and maternal age as the most significant predictors of iron supplementation adherence. This predictor hierarchy, particularly the prominence of socioeconomic factors (wealth status) and healthcare access (ANC visits) aligns with findings from Ethiopia, where similar machine learning approaches identified comparable determinants of micronutrient behaviors among vulnerable populations (43).

In this study, maternal age was identified as a significant factor associated with the uptake of iron supplementation. Notably, women aged 15–19 indicated a greater likelihood of adhering to iron supplementation during pregnancy compared to their older counterparts. This is consistent with findings from previous studies conducted from Pakistan (33), Ethiopia (44), and sub-Saharan African countries (20) that is similarly reported younger maternal age were positively associated with higher adherence to antenatal iron supplementation.

Our study found that women with higher levels of education were more likely to use iron supplements during pregnancy compared to those with no formal education. This finding is consistent with studies conducted

in Pakistan, East Africa, and Ethiopia, which have identified maternal education as a key determinant of adherence to iron and folic acid supplementation (33,45–47). A possible explanation for this association is that education enhances women's health literacy, increasing their awareness of the benefits of micronutrient supplementation during pregnancy. Educated women are also more likely to access antenatal care and better understand the maternal and child health implications of iron deficiency.

Employment status emerged as a significant factor influencing iron supplementation uptake among pregnant women. The findings of this study indicate that employed women were more likely to adhere to iron supplementation recommendations compared to their unemployed counterparts. This association may be attributed to the fact that employment can enhance women's financial autonomy and access to health-related information, which in turn promotes positive health-seeking behaviors. These results are consistent with prior study conducted in Pakistan that identified employment as a predictor of better utilization of antenatal and nutritional services during pregnancy (33).

Furthermore, the present study identified a significant association between the number of antenatal care (ANC) visits and adherence to iron supplementation among pregnant women. Specifically, women who attended more than four ANC visits were more likely to adhere to iron supplement recommendations compared to those with fewer than four visits. This finding aligns with previous research conducted across various countries, including Ethiopia, Malawi, sub-Saharan African nations, Pakistan, and India, which consistently reported that increased frequency of ANC visits is positively associated with better adherence to iron supplementation (13,20,33,38,48,49). These studies suggest that frequent ANC visits enhance maternal awareness, improve access to supplements, and strengthen the health provider–client relationship, thereby facilitating adherence to iron intake during pregnancy.

Moreover, media exposure was found to be positively associated with iron supplementation uptake among pregnant women. Similarly, this finding is supported by prior studies conducted in Indonesia, Ethiopia, and various low- and lower-middle-income Asian countries, which indicated that women who were regularly exposed to media sources such as television or radio, were more likely to adopt recommended maternal health practices (12,15,50). Therefore, media plays a critical role in raising awareness by disseminating essential information regarding the benefits of iron supplementation, which in turn promotes positive health-seeking behaviors and adherence to antenatal care guidelines.

Additionally, the wealth index of the respondents was strongly associated with iron supplement uptake. The findings indicated that pregnant women with a higher wealth index were more likely to use iron supplements than their poorer counterparts. This suggests that individuals with higher socioeconomic status have greater access to healthcare services and are more likely to afford and adhere to recommended supplementation during pregnancy. These findings are supported by similar studies conducted in Malawi (48), Ethiopia (51), Pakistan (33), and 22 sub-Saharan African countries (13).

The region was found to be significantly associated with iron supplementation among pregnant women, indicating that geographic location plays a critical role in maternal health service utilization. The study findings revealed notable geographic variations in the uptake of iron supplementation across different regions. This finding aligns with previous studies conducted in Pakistan (33) that reported regional disparities in maternal health service utilization are play a vital role in the uptake of iron supplements.

5. Conclusion and recommendation

This study demonstrated that iron supplement use among pregnant women in Somaliland remains suboptimal, with only 28.8% reporting adherence. Using data from the 2020 Somaliland Health and Demographic Survey and a variety of machine learning models, particularly the Random Forest model demonstrated the strongest predictive capability, key predictors of iron supplement uptake were identified most notably wealth status, region, antenatal care (ANC) visits, and maternal age. These findings underscore the importance of socioeconomic conditions, geographical location, and healthcare access in influencing iron supplementation uptake. Targeted interventions, particularly those addressing wealth disparities and promoting antenatal care, are therefore essential to improve maternal and child health outcomes in Somaliland.

6. Strength and limitations

Our analysis was based a nationally representative dataset, the 2020 Somaliland DHS, and employing advanced machine learning techniques to identify key predictors of iron supplementation, which is the strength of this study. However, a limitation of the study is it used a cross-sectional design, which limits the establishment of causal relationships between predictors (independent variables) and iron supplementation uptake. Furthermore, the study used secondary data for analysis. The data have missing value in the variable and some variables are not found in the dataset which are some important variables missing from the analysis. Therefore, missingness in the data and variables were the main limitation of this study.

Abbreviations

AUC	–	Area Under the Curve
ANC	–	Antenatal Care
CI	–	Confidence Interval
DT	–	Decision Tree
EAs	–	Enumeration Areas
FN	–	False Negative
FP	–	False Positive
GBM	–	Gradient Boosting Machine
GIS	–	Geographic Information System
GBM	–	Gradient Boosting Machine
KNN	–	K-Nearest Neighbors
LR	–	Logistic Regression
PPS	–	Probability Proportional to Size
RF	–	Random Forest
ROC	–	Receiver Operating Characteristic
SVM	–	Support Vector Machine
SLHDS	–	Somaliland Health and Demographic Survey
SMOTE	–	Synthetic Minority Over-sampling Technique
TP	–	True Positive
TN	–	True Negative
WHO	–	World Health Organization

References

1. Stevens GA, Finucane MM, De-Regil LM, Paciorek CJ, Flaxman SR, Branca F, et al. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995–2011: a systematic analysis of population-representative data. Lancet Glob Heal. 2013;1(1):e16–25.
2. WHO C. WHO global database on anemia. 2008. Worldw Preval anemia. 1993;2005.
3. Yekta Z, AYAT EH, Pourali R, Farzin A. Predicting factors in iron supplement intake among pregnant women in urban care setting. 2008;
4. Bothwell TH. Iron requirements in pregnancy and strategies to meet them. Am J Clin Nutr. 2000;72(1):257S-264S.
5. Trumbo P, Yates AA, Schlicker S, Poos M. Dietary Reference Intakes. J Am Diet Assoc. 2001;101(3):294.
6. Stoltzfus RJ, Dreyfuss ML. Guidelines for the use of iron supplements to prevent and treat iron deficiency anemia. Vol. 2. Ilsi Press Washington, DC; 1998.
7. Allen LH. Anemia and iron deficiency: effects on pregnancy outcome. Am J Clin Nutr. 2000;71(5):1280S-1284S.
8. Milman N. Oral iron prophylaxis in pregnancy: not too little and not too much! J Pregnancy. 2012;2012(1):514345.
9. Menon KC, Ferguson EL, Thomson CD, Gray AR, Zodpey S, Saraf A, et al. Effects of anemia at different stages of gestation on infant outcomes. Nutrition. 2016;32(1):61–5.

10. Organization WH. Daily iron and folic acid supplementation in pregnant women: guideline. In: Daily iron and folic acid supplementation in pregnant women: guideline. 2012. p. 27.
11. Garzon S, Cacciato PM, Certelli C, Salvaggio C, Magliarditi M, Rizzo G. Iron deficiency anemia in pregnancy: Novel approaches for an old problem. Oman Med J. 2020;35(5):e166.
12. Efendi F, Israfil I, Ramadhan K, McKenna L, Alem AZ, Malini H. Factors associated with receiving iron supplements during pregnancy among women in Indonesia. Electron J Gen Med. 2023;20(5).
13. Ba DM, Ssentongo P, Kjerulff KH, Na M, Liu G, Gao X, et al. Adherence to iron supplementation in 22 sub-Saharan African countries and associated factors among pregnant women: a large population-based study. Curr Dev Nutr. 2019;3(12):nzz120.
14. Ogundipe O, Hoyo C, Østbye T, Oneko O, Manongi R, Lie RT, et al. Factors associated with prenatal folic acid and iron supplementation among 21,889 pregnant women in Northern Tanzania: a cross-sectional hospital-based study. BMC Public Health. 2012;12:1–10.
15. Agegnehu CD, Tesema GA, Teshale AB, Alem AZ, Yeshaw Y, Kebede SA, et al. Spatial distribution and determinants of iron supplementation among pregnant women in Ethiopia: a spatial and multilevel analysis. Arch Public Heal. 2021;79(1):143.
16. Kibret KT, Chojenta C, D'Arcy E, Loxton D. Spatial distribution and determinant factors of anaemia among women of reproductive age in Ethiopia: a multilevel and spatial analysis. BMJ Open. 2019;9(4):e027276.
17. Conceição P. Human development report 2020-the next frontier: Human development and the anthropocene. United Nations Dev Program Hum Dev Rep. 2020;
18. Staff U. The state of the world's children 2011-executive summary: Adolescence an age of opportunity. Unicef; 2011.
19. Group WBIEDDD, Group WBIEDDD. World development indicators. World Bank; 1978.
20. Zegeye B, Adjei NK, Olorunsaiye CZ, Ahinkorah BO, Ameyaw EK, Seidu AA, et al. Pregnant women's decision-making capacity and adherence to iron supplementation in sub-Saharan Africa: a multi-country analysis of 25 countries. BMC Pregnancy Childbirth. 2021;21:1–10.
21. Setia MS. Methodology series module 3: Cross-sectional studies. Indian J Dermatol. 2016;61(3):261–4.
22. Organization WH. Guideline: daily iron and folic acid supplementation in pregnant women. In: Guideline: Daily iron and folic acid supplementation in pregnant women. 2012. p. 32.
23. UIf S. International standard classification of education: ISCED 2011. Montr UIS. 2012;85.
24. Organization WH. WHO recommendations on antenatal care for a positive pregnancy experience. World Health Organization; 2016.
25. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Health Policy Plan. 2006;21(6):459–68.
26. Chawla N V, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.
27. Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J. 2020;20(1):3–29.
28. Machado MR, Karray S, De Sousa IT. LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In: 2019 14th international conference on computer science & education (ICCSE). IEEE; 2019. p. 1111–6.
29. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016. p. 785–94.
30. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics. 2018;15(1):41–51.
31. Sathyanarayanan S, Tantri BR. Confusion matrix-based performance evaluation metrics. African J Biomed Res. 2024;4023–31.
32. Boti N, Bekele T, Godana W, Getahun E, Gebremeskel F, Tsegaye B, et al. Adherence to Iron-Folate Supplementation and Associated Factors among Pastoralist's Pregnant Women in Burji Districts, Segen Area People's Zone, Southern Ethiopia: Community-Based Cross-Sectional Study. Int J Reprod Med. 2018;2018(1):2365362.
33. Nisar Y Bin, Dibley MJ, Mir AM. Factors associated with non-use of antenatal iron and folic acid supplements among Pakistani women: a cross sectional household survey. BMC Pregnancy Childbirth.

2014;14:1–12.
34. Ali SA, Ali SA, Razzaq S, Khowaja N, Gutkind S, Raheman FU, et al. Predictors of iron consumption for at least 90 days during pregnancy: Findings from National Demographic Health Survey, Pakistan (2017–2018). BMC Pregnancy Childbirth. 2021;21(1):352.
35. Knudsen VK, Hansen HS, Ovesen L, Mikkelsen TB, Olsen SF. Iron supplement use among Danish pregnant women. Public Health Nutr. 2007;10(10):1104–10.
36. Lutsey PL, Dawe D, Villate E, Valencia S, Lopez O. Iron supplementation compliance among pregnant women in Bicol, Philippines. Public Health Nutr. 2008;11(1):76–82.
37. Kiwanuka TS, Ononge S, Kiondo P, Namusoke F. Adherence to iron supplements among women receiving antenatal care at Mulago National Referral Hospital, Uganda-cross-sectional study. BMC Res Notes. 2017;10:1–6.
38. Haile D, Tabar L, Lakew Y. Differences in spatial distributions of iron supplementation use among pregnant women and associated factors in Ethiopia: evidence from the 2011 national population based survey. BMC Pregnancy Childbirth. 2017;17:1–8.
39. Ngusie HS, Enyew EB, Walle AD, Tilahun Assaye B, Kasaye MD, Tesfa GA, et al. Employing machine learning techniques for prediction of micronutrient supplementation status during pregnancy in East African Countries. Sci Rep. 2024;14(1):23827.
40. Taye EA, Woubet EY, Hailie GY, Arage FG, Zerihun TE, Zegeye AT, et al. 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. 2025;25(1):901.
41. Khan Majlish MR, Tawhid SA, Kibria G, Sakib N. Predicting Unintended Pregnancy in Senegal by using Machine Learning Models: Evidence from Senegal DHS 2023. medRxiv. 2024;2011–24.
42. Naznin S, Uddin MJ, Kabir A. Identifying determinants of under-5 mortality in Bangladesh: A machine learning approach with BDHS 2022 data. PLoS One. 2025;20(6):e0324825.
43. Zemariam AB, Adisu MA, Habesse AA, Abate BB, Bizuayehu MA, Wondie WT, et al. Employing advanced supervised machine learning approaches for predicting micronutrient intake status among children aged 6–23 months in Ethiopia. Front Nutr. 2024;11:1397399.
44. Gebremichael TG, Welesamuel TG. Adherence to iron-folic acid supplement and associated factors among antenatal care attending pregnant mothers in governmental health institutions of Adwa town, Tigray, Ethiopia: Cross-sectional study. PLoS One. 2020;15(1):e0227090.
45. Alemu NL, Roba KT, Getaneh HD, Raru TB, Daba AK, Worku N, et al. Adherence to Iron and Folic Acid Supplementation and its determinants among pregnant women in East Africa: Analysis of Demographic and Health Surveys data from nine East African countries. PLoS One. 2025;20(7):e0327410.
46. Mulugeta SS. Geographical disparities and determinants of adherence to iron folate supplementation among pregnant women in Ethiopia: spatial and multilevel analysis of the Ethiopian Mini Demographic and Health Survey of 2019. BMJ Open. 2022 Sep;12(9):e061900.
47. Asmamaw DB, Debebe Negash W, Bitew DA, Belachew TB. Poor adherence to iron-folic acid supplementation and associated factors among pregnant women who had at least four antenatal care in Ethiopia. A community-based cross-sectional study. Front Nutr. 2022;9:1023046.
48. Titilayo A, Palamuleni ME, Omisakin O. Sociodemographic factors influencing adherence to antenatal iron supplementation recommendations among pregnant women in Malawi: Analysis of data from the 2010 Malawi Demographic and Health Survey. Malawi Med J. 2016;28(1):1–5.
49. Roy MP, Mohan U, Singh SK, Singh VK, Srivastava AK. Socio-economic determinants of adherence to iron and folic acid tablets among rural ante-natal mothers in Lucknow, India. Natl J Community Med. 2013;4(3):386–91.
50. Warvadekar K, Reddy JC, Sharma S, Dearden KA, Raut MK. Socio-demographic and economic determinants of adherence to iron intake among pregnant women in selected low and lower middle income countries in Asia: insights from a cross-country analyses of global demographic and health surveys. Intl J Comm Med Public Heal. 2018;5(4):1552–69.
51. Agegnehu G, Atenafu A, Dagne H, Dagnew B. Adherence to iron and folic acid supplement and its

associated factors among antenatal care attendant mothers in Lay Armachiho health centers, Northwest, Ethiopia, 2017. Int J Reprod Med. 2019;2019(1):5863737.

Authorship contribution statement

Abdifatah Ibrahim Mouse: Topic Selection, Conceptualization, Data curation, Data management, Methodology, Software, Formal analysis, Discussion, and Writing – final draft.
Omran Salih: Validation, Supervision and Writing – review & editing, and approved the submitted version.
Abdisalam Hassan Muse: Validation, Supervision and Writing – review & editing, and approved the submitted version.

Declarations

Ethics approval and consent to participate

Permission was granted to download and use the data from http://www.dhs.program.com before conducting the study. Ethical clearance was obtained from the Institution Review Board of the DHS Program, ICF International. The procedures for DHS public-use data sets were approved by the Institution Review Board. Identifiers for respondents, households, or sample communities were not allowed in any way, and the names of individuals or household addresses were not included in the data files. The number for each EA in the data file does not have labels to show their names or locations. There were no patients or members of the public involved since this study used a publicly available data set

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics, consent to participate, and consent to publish

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable. This study used secondary publicly available data, and no direct involvement of human participants was required.

Competing interest

The authors have no competing interests to declare that are relevant to the content of this article. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The dataset used in this study is publicly accessible on the Somalia National Bureau of Statistics website at https://microdata.nbs.gov.so/index.php/catalog/50.

Acknowledgments

The authors are grateful to the Demographic and Health Surveys Program for granting them access to the dataset.

Author details

[1]Faculty of Science and Humanities, School of Postgraduate Studies and Research (SPGSR), Amoud University, Hargeisa 25263, Somaliland
[2]Institute of Systems Science, Durban University of Technology, Durban 4000, South Africa
[3]Research and Innovation Center, Amoud University, Borama 25263, Somaliland