

Machine Learning Prediction of Iron Supplement Utilization Among Pregnant Women in Somaliland: Evidence from the Somaliland Demographic and Health Survey 2020

Hamda Jama Yousuf

hamda@ogaansho.org

Ogaansho Research and Consultancy Centre

Abdiasis Adem Omar

Ogaansho Research and Consultancy Centre

Suhur A. Ahmed

Ogaansho Research and Consultancy Centre

Ahmed Abdi Omar

Ogaansho Research and Consultancy Centre

Ahmed Abdirahman Farah

Ogaansho Research and Consultancy Centre

Research Article

Keywords: Iron supplementation, pregnancy, maternal health, machine learning, antenatal care, Somaliland

Posted Date: March 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9082298/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Machine Learning Prediction of Iron Supplement Utilization Among Pregnant Women in Somaliland: Evidence from the Somaliland Demographic and Health Survey 2020

Hamda Jama Yousuf¹, Abdiasis Adem Omer¹, Suhur A. Ahmed¹, Ahmed Abdi Omar¹, Ahmed Abdirahman Farah¹

¹ Ogaansho Research and Consultancy Center, Burao, Somaliland

^{1*} Corresponding Author: Hamda Jama Yousuf (hamda@ogaansho.org)

Abstract

Background: Iron supplementation during pregnancy is a key public health intervention for preventing maternal anemia and improving maternal and neonatal outcomes. This study aimed to identify the determinants and predict iron supplement utilization among pregnant women in Somaliland using machine learning techniques.

Methods: This study used data from the 2020 Somaliland Demographic and Health Survey (SLDHS), a nationally representative cross-sectional survey. The outcome variable was iron supplement utilization during pregnancy. Descriptive statistics and bivariate analyses were conducted to examine associations between independent variables and iron supplement use. Several supervised machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, and Extreme Gradient Boosting (XGBoost), were applied to predict iron supplement utilization. The dataset was split into training (80%) and testing (20%) subsets, and model performance was evaluated using accuracy, precision, recall, specificity, and area under the receiver operating characteristic curve (AUC-ROC).

Results: The prevalence of iron supplement utilization among pregnant women was 27%, indicating low coverage in the study population. Bivariate analysis revealed that region, educational level, wealth index, distance to a health facility, husband's employment status, antenatal care (ANC) visits, and media exposure were significantly associated with iron supplement use ($p < 0.05$). Among the machine learning models, Support Vector Machine achieved the highest accuracy (82.6%), followed by Logistic Regression (81.7%) and Random Forest (80.6%). Logistic Regression (AUC = 0.853), Random Forest (AUC = 0.852), and SVM (AUC = 0.850) demonstrated the strongest discriminatory performance. Feature importance analysis indicated that ANC utilization, husband's employment status, media exposure, and distance to health facilities were the most influential predictors of iron supplement utilization.

Conclusion: Iron supplementation during pregnancy remains substantially low in Somaliland. Maternal healthcare utilization, socioeconomic status, and access to health information play important roles in determining iron supplement use. Machine learning approaches demonstrated strong predictive performance in identifying key determinants of iron supplementation. Strengthening antenatal care services, improving maternal health education through mass media, and addressing geographic barriers to healthcare access may significantly improve iron supplementation coverage and maternal nutrition in Somaliland.

Keywords: Iron supplementation; pregnancy; maternal health; machine learning; antenatal care; Somaliland

Background

Iron and folic acid are essential micronutrients for normal physiological function, growth, and development, as well as maintenance of life. Like many other nutrients, their demand increases during pregnancy, creating a need for supplementation to meet the daily requirements in pregnancy. The high nutrient demand during pregnancy may not be met by a regular diet because of insufficient amounts and/or low bioavailability, especially in developing countries. Due to this disproportion, deficiency in iron or folic acid causes an imbalance between demand and supply, resulting in anemia. Iron and folic acid and other micronutrient deficiencies are highly prevalent in Kenya, particularly at crucial stages of the life cycle, such as pregnancy, which can negatively impact the health of both mothers and children, before, during, and after birth(1).

All over the world, anemia is one of the public health problems and continues to be a universal top cause of frailty and the highest critical global health matter. This is because in a pregnancy, it is extremely predominant in both industrialized and non-industrialized countries. Recent evidence from the World Health Organization (WHO) document revealed that nearly 38% (32 million) of women in pregnancy are anemic in the world. Out of this, 46.3% (9.2 million) of them are in Africa. However, the elucidations of the rate mostly show a discrepancy in the world from place to place. For instance, there is substantial deviation in the rate of anemia in pregnancy within advanced countries like the USA, in which the rate is 18%, in Australia, 20%, in Singapore 67.8% and in China, 70%; whereas the rates rise through trimesters. But the magnitude of the rate is getting higher in unindustrialized nations; for example, in Ethiopia, 50.1%, in Sudan, 53%, in Guinea 71% and in Pakistan, 76.7%. These are the fundamental rationale problems related to anemia, which is one of the basic concerns of public health issues in the world in general and in Africa in particular(2).

Anemia is a global public health problem affecting two billion people worldwide. Particularly pregnant women and preschool-age children take its disproportionate burden. Globally, almost half of all preschool children (47.4%) and pregnant women (41.8%), and close to one-third of non-pregnant women (30.2%) are anemic. Though anemia has multifaceted causes, half of its burden is attributed to Iron Deficiency. Pregnancy-related anemia remains the top among the leading causes of global health burden, with more than half attributed to iron deficiency, the most common form of micronutrient malnutrition, and a prevalent nutritional deficiency leading to significant morbidity and mortality globally. The global prevalence of pregnancy-related anemia spans from 41.8 to 43.8%, translating to approximately 59 million pregnant women, with the greatest (61.3%) burden in Africa, and South East Asia at 52.5% (3).

Iron and folic acid supplementation (IFAS) during pregnancy has been recommended as one of the strategies to address this public health problem. It is an affordable and effective global strategy for prevention and control of anemia during pregnancy, resulting in reduced maternal-child morbidity and mortality(1). Folate deficiency at conception and in early pregnancy is associated with increased risk of neural tube defects. In both developing and developed countries,

supplementation with 400 micrograms per day ($\mu\text{g/d}$) of folic acid is recommended for all women of childbearing age. However, the prevalence of anemia is increasing in LMICs, with a rate of 35.6–76.7%. Africa has the highest prevalence of anemia, with 57% of pregnant women. Several studies carried out in Ethiopia also revealed that the prevalence of anemia among pregnant women ranges from 21 to 54%. Maternal anemia is associated with mortality and morbidity for the mother and baby, including the risk of miscarriages, stillbirths, prematurity, and low birth weight. It impairs children's development and learning, too, further impacting economic productivity and development(4). In developing countries, every second pregnant woman is estimated to be anemic. In Kenya, despite efforts to control and prevent anemia in pregnancy, this public health problem persists, with a national prevalence of 55.1%, resulting in estimated 10% maternal and 20% perinatal deaths, respectively. The prevalence is similar to that found in Asia, where the problem is the second-highest cause of maternal mortality.

By 2025, the World Health Organization (WHO) aims to reduce anemia in women of reproductive age by half. In Ethiopia, nutrition is integrated into the health sector transformation plan in the form of micronutrient interventions to prevent the occurrence of anemia and improve the nutritional condition of mothers during pregnancy. Maternal and child mortality remain high in Ethiopia, with 412 maternal deaths per 100,000 live births and 67 child deaths per 103 live births reported. Taking this into account, the Ethiopian government has demonstrated its commitment to averting nutritional problems by adopting food and nutrition policies, strategies, and programs. In addition, the government designed the Seqota declaration to eliminate stunting by 2030. Various factors, such as socio-demographic and health factors, determine the compliance with iron supplementation among pregnant women. All pregnant women in places where anemia is common should take iron supplements, according to the WHO. Despite the WHO's recommendation, iron supplementation is nevertheless uncommon in many countries, particularly in those with limited resources(5).

In Somaliland, maternal anemia and micronutrient deficiencies remain important public health concerns that can negatively affect both maternal and neonatal health outcomes. Although iron supplementation is recommended as part of routine antenatal care services, evidence regarding the utilization of iron supplements among pregnant women in Somaliland is limited. Nationally representative data from the Somaliland Demographic and Health Survey (SLDHS) provide an opportunity to better understand the patterns and determinants of iron supplement utilization among pregnant women. Identifying the factors influencing iron supplement use is essential for improving maternal nutrition programs and strengthening maternal health services in the country.

Despite global recommendations and ongoing public health efforts to promote iron and folic acid supplementation during pregnancy, the utilization of iron supplements remains suboptimal in many low- and middle-income countries. Poor adherence to iron supplementation continues to contribute to the high burden of maternal anemia and its associated complications, including increased risk of maternal mortality, preterm birth, and low birth weight. In many settings, including sub-Saharan Africa, limited access to health services, inadequate awareness, and socioeconomic barriers continue to hinder optimal use of iron supplementation among

pregnant women. However, evidence on the determinants of iron supplement utilization among pregnant women in Somaliland remains scarce. Understanding these determinants is critical for designing targeted interventions aimed at improving maternal nutrition and reducing anemia-related complications during pregnancy. This study contributes to the achievement of the Sustainable Development Goals (SDGs), particularly SDG 3 (Good Health and Well-being), which aims to reduce maternal mortality and improve maternal health outcomes. By examining the determinants and predicting the utilization of iron supplementation among pregnant women in Somaliland, the findings of this study can support evidence-based interventions aimed at reducing maternal anemia and improving pregnancy outcomes. In addition, the study indirectly supports SDG 2 (Zero Hunger), which focuses on ending all forms of malnutrition, including micronutrient deficiencies among women of reproductive age. To our knowledge, this is the first study applying machine learning algorithms to predict iron supplementation utilization among pregnant women in Somaliland using nationally representative survey data.

Methods and materials

Study setting

Somaliland, a self-proclaimed sovereign territory in the Horn of Africa, covers over 176,000 square kilometers and shares borders with Somalia, Ethiopia, and Djibouti. Its 4.5 million residents are largely of Somali ethnicity and the Islamic faith, with a significant portion of the population living in the capital city of Hargeisa. The landscape varies from an extensive 850 km coast to northern mountainous terrain. Administratively, the area is divided into 22 districts across six provinces: Awdal, Marodijeh, Sahil, Togdheer, Sool, and Sanaag. Despite being a major exporter of livestock, Somaliland's economic progress is restricted by limited external investment and its unrecognized status on the global stage(6).

Study Design, Data Source, and Study Period

The research utilized data from the 2020 Somaliland Demographic Health Survey (SLHDS), the first nationally representative health survey in the country. This cross-sectional survey was conducted from August 2018 to December 2019. This design is appropriate for examining the determinants and predictors of iron supplement utilization among pregnant women in Somaliland. The survey aimed to strengthen national health data systems and support evidence-based planning in Somaliland (6).

Study Variables

The outcome variable

The outcome variable in this study was iron supplement utilization during pregnancy. This variable was derived from the survey question asking whether women took iron tablets or syrup during their most recent pregnancy. The outcome variable was coded as a binary variable, where 1 indicated iron supplement use and 0 indicated non-use.

Independent variables

The independent variables comprised a set of explanatory factors selected based on their theoretical relevance and evidence from previous studies on maternal health service utilization. These variables included both sociodemographic and reproductive health characteristics. The

sociodemographic variables consisted of maternal age, region, place of residence, educational level, household wealth index, media exposure, husband's employment status, and respondent employment status. In addition, household characteristics such as family size were included. The reproductive and healthcare-related variables comprised parity, age at first birth, antenatal care (ANC) visits during pregnancy, and perceived distance to a health facility. These variables were selected to examine their association with iron supplement utilization and to identify the most influential predictors through both statistical and machine learning analyses.

Data preprocessing

Data processing and analysis were conducted using both statistical and machine learning techniques. Before analysis, the dataset was cleaned and examined for completeness, consistency, and potential missing values. Relevant variables were extracted from the Somaliland Demographic and Health Survey (SLDHS) 2020 dataset, and categorical variables were appropriately recoded to facilitate analysis. Sampling weights provided in the DHS dataset were applied to ensure that the results were representative of the national population.

Statistical analysis

Descriptive statistics, including weighted frequencies and percentages, were computed to summarize the characteristics of the study participants. Bivariate associations between the independent variables and iron supplement utilization were examined using the Chi-square test, with the application of survey weights to ensure that the findings were representative of the target population. A p-value of less than 0.05 was considered statistically significant. Variables identified as significant in the bivariable analysis were considered in the machine learning phase. Feature importance was derived from the trained machine learning models to identify the most influential predictors of iron supplement utilization. All statistical analyses were performed using Stata (Version 17) and Python (Version 3.9), with relevant analytical libraries used for data management, model development, and performance evaluation.

Machine learning models

Given the nature of the dataset, which contains complex relationships among sociodemographic, reproductive, and healthcare-related variables, and the research objective of predicting iron supplement utilization, several supervised machine learning algorithms were implemented. These algorithms were selected because of their robustness, interpretability, and effectiveness in binary classification tasks. The models used in this study included Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, and Extreme Gradient Boosting (XGBoost). Each algorithm offers a distinct approach for modeling relationships between predictors and the outcome variable.

1. Logistic Regression:

Logistic regression is a widely used statistical model for binary classification. It estimates the probability of an outcome using the logistic function, which constrains predictions between 0 and 1. The model provides interpretable coefficients that help assess the relationship between independent variables and iron supplement utilization.(7, 8)

2. Decision Tree:

A decision tree is a non-parametric model that partitions the dataset into subsets based on the values of predictor variables. It creates a hierarchical tree-like structure where each node represents a decision rule. This method is highly interpretable and capable of capturing complex and non-linear relationships among variables(9).

3. Random Forest:

Random Forest is an ensemble learning method that constructs multiple decision trees using random subsets of the data and predictor variables. The final classification is obtained through majority voting among the individual trees. This approach improves predictive accuracy and reduces the risk of overfitting while also providing feature importance measures(10).

4. Support Vector Machine (SVM):

Support Vector Machine is a powerful classification algorithm that identifies an optimal hyperplane separating observations belonging to different classes in a high-dimensional feature space. By maximizing the margin between classes, SVM improves model generalization and classification performance(11, 12).

5. K-Nearest Neighbors (KNN):

KNN is a non-parametric algorithm that classifies observations based on the majority class among the k closest neighbors in the feature space. Distance metrics such as Euclidean distance are used to determine similarity between observations(13).

6. Naïve Bayes:

Naïve Bayes is a probabilistic classifier based on Bayes' theorem. It assumes conditional independence between predictors given the outcome class. Despite this assumption, it often performs well in classification tasks and provides efficient and fast predictions(14).

7. Extreme Gradient Boosting (XGBoost):

XGBoost is an advanced ensemble learning algorithm based on gradient boosting. It builds models sequentially, where each new tree attempts to correct errors made by the previous model. XGBoost is widely recognized for its high predictive accuracy and efficiency in handling complex datasets(9, 15).

Model development and training

After completing data preprocessing, the dataset was prepared for machine learning model development. The final dataset was randomly divided into training and testing subsets, where 80% of the data were used for training the models, and 20% were reserved for testing and performance evaluation. The training dataset was used to build and optimize the machine learning models, while the testing dataset provided an unbiased assessment of model performance on previously unseen data. To improve model robustness and reduce the risk of overfitting, a 10-fold cross-validation technique was applied during the training phase. This method allowed the models to be trained and validated across multiple subsets of the training data, thereby enhancing the reliability and generalizability of the predictive results. Each machine learning algorithm was trained using the selected predictor variables to classify iron supplement utilization (use vs. non-use). During the

training process, model parameters were optimized to achieve the best predictive performance. The trained models were then applied to the testing dataset to evaluate their classification performance using several evaluation metrics, including accuracy, precision, recall, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). This modeling framework enabled a systematic comparison of the predictive performance of different machine learning algorithms and helped identify the most reliable model for predicting iron supplement utilization.

Model evaluation

To assess the predictive performance of the developed machine learning (ML) models on the held-out testing dataset, several evaluation metrics were employed. These metrics included accuracy, precision, recall (sensitivity), specificity, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC).

Accuracy represents the proportion of correctly classified observations among the total number of cases examined. It is calculated

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

where:

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

Precision measures the proportion of correctly predicted positive cases out of all instances classified as positive:

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

Recall (Sensitivity) quantifies the model's ability to correctly identify true positive cases:

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

Specificity measures the ability of the model to correctly identify true negative cases:

$$Specificity = TN / (TN + FP)$$

The F1-score is the harmonic mean of precision and recall, providing a balanced measure that accounts for both false positives and false negatives:

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall)$$

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was also used as a comprehensive indicator of model performance. The ROC curve plots the true positive rate (sensitivity) against the false positive rate across different classification thresholds. A higher AUC-ROC value indicates better discriminatory ability of the model in distinguishing between iron supplement users and non-users.

Additionally, confusion matrices and classification reports were generated to provide a detailed understanding of the models' predictive performance. Feature importance was extracted from the trained models using coefficients for Logistic Regression, feature importance measures

for Decision Tree and Random Forest models, and permutation importance for SVM to identify the key determinants influencing iron supplement utilization.

Result

Prevalence of Iron Supplement Utilization

Figure 1 presents the prevalence of iron supplement use among the study participants. The results indicate that a substantial majority of women (**73%**) did not use iron supplements during pregnancy, whereas only **27%** reported using iron supplements. This finding demonstrates a relatively low level of iron supplementation among pregnant women in the study population. Given that iron supplementation is a critical intervention for preventing maternal anemia and improving maternal and fetal health outcomes, the low utilization observed in this study may reflect gaps in maternal health service utilization

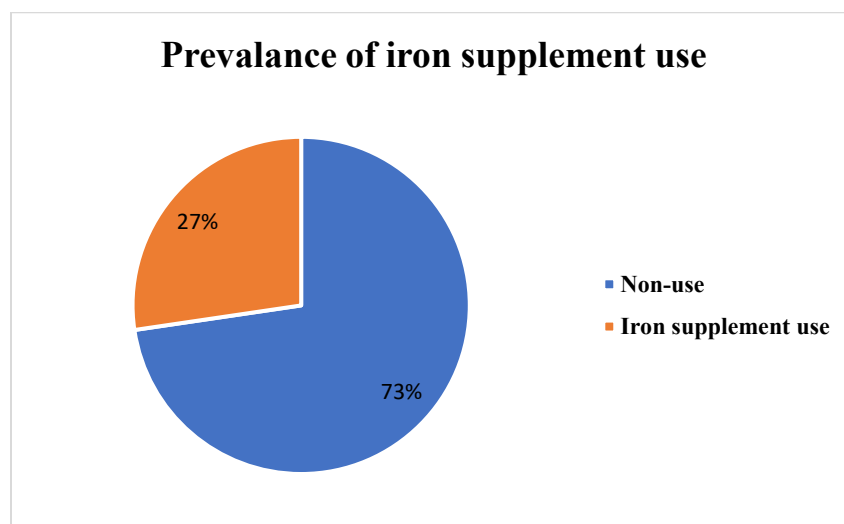


Figure 1: Proportion of iron supplement use

Sociodemographic Characteristics of the Study Participants

Table 1 presents the weighted descriptive characteristics of the study participants. The largest proportion of women was aged 25–29 years (29.09%), followed by those aged 30–34 years (23.14%) and 35–39 years (16.99%), while the smallest proportion was observed among women aged 45–49 years (2.05%). Regarding regional distribution, the highest proportion of respondents was from Marodijeh (27.25%) and Togdheer (25.47%), whereas Sahil had the lowest representation (5.17%). In terms of residence, more than half of the women lived in rural areas (57.17%), while 42.83% resided in urban areas. Educational attainment was generally low among the respondents, with the majority having no formal education (78.29%), followed by primary education (15.36%), secondary education (4.00%), and higher education (2.35%). Most households had four or more family members (79.24%), whereas 20.76% had fewer than four members. With regard to household economic status, 29.78% of respondents were in the highest wealth quintile, followed by 28.21% in the lowest quintile, indicating variation in socioeconomic status among the participants. Concerning reproductive characteristics, 44.60% of women were

multiparous and 42.67% were grand multiparous, while only 12.72% were primiparous. Nearly half of the respondents (43.56%) had their first birth before the age of 20 years, while only 2.42% had their first birth at age 30 years or older. Access to healthcare services was reported as a challenge for many women, with 61.85% indicating that distance to a health facility was a big problem. In terms of employment status, almost all respondents (99.40%) reported that they had not worked in the past 12 months, whereas about half (49.75%) indicated that their husbands had worked during the same period. Regarding maternal health service utilization, more than half of the women (52.56%) did not utilize antenatal care (ANC) services, 27.97% partially utilized ANC, and only 19.47% adequately utilized ANC services. Additionally, the majority of respondents (61.52%) reported no media exposure, while 38.48% had some level of exposure to media.

Table 1: Descriptive Characteristics of the Study

Characteristic	Category	Weighted frequency	Weighted percentage (%)
Age	15–19	202.0	4.97
	20–24	665.6	16.38
	25–29	1182.5	29.09
	30–34	940.7	23.14
	35–39	690.6	16.99
	40–44	300.0	7.38
	45–49	83.3	2.05
Region	Awdal	382.9	9.42
	Marodijeh	1107.0	27.25
	Sahil	210.2	5.17
	Togdheer	1035.6	25.47
	Sool	583.9	14.36
	Sanaag	746.4	18.34
Place of residence	Rural	2324.6	57.17
	Urban	1741.4	42.83
Education level	No education	3183.3	78.29
	Primary	624.5	15.36
	Secondary	162.6	4.00
	Higher	95.6	2.35
Family size	<4 members	844.1	20.76
	≥4 members	3221.9	79.24
Wealth index	Lowest	1147.0	28.21
	Second	557.2	13.71
	Middle	421.2	10.36
	Fourth	729.3	17.94
	Highest	1211.3	29.78
Parity	Primipara	517.3	12.72
	Multipara	1813.4	44.60
	Grand multipara	1735.3	42.67

Age at first birth	<20	1770.7	43.56
	20–24	1553.9	38.19
	25–30	643.7	15.83
	≥30	98.4	2.42
Distance to health facility	Big problem	2513.8	61.85
	Not a big problem	1552.2	38.15
Husband worked last 12 months	Yes	2023.6	49.75
	No	2013.3	49.50
	Don't know	30.1	0.75
Respondent worked last 12 months	Yes	24.4	0.60
	No	4041.6	99.40
ANC visits	Not utilized	2136.7	52.56
	Partially utilized	1137.4	27.97
	Adequately utilized	792.0	19.47
Media exposure	No	2500.0	61.52
	Yes	1566.0	38.48

Bivariate Analysis of Factors Associated with Iron Supplement Utilization

Table 2 presents the results of the bivariate analysis examining the association between selected predictors and iron supplement use among women. The findings indicate that region, education level, wealth index, distance to a health facility, husband's employment status, antenatal care (ANC) visits, and media exposure were significantly associated with iron supplement use ($p < 0.05$). Iron supplement use varied significantly across regions ($p < 0.001$), with higher proportions observed among women from Marodijeh (35.82%) and Togdheer (34.47%), while lower proportions were reported in Sool (5.91%) and Sanaag (8.95%). Educational attainment also showed a strong association with iron supplement use ($p < 0.001$), as women with secondary (7.79%) and higher education (4.34%) were more likely to use iron supplements compared to those with no formal education (66.01%) among users.

Household wealth status demonstrated a significant relationship with iron supplement utilization ($p < 0.001$). Women in the highest wealth quintile (46.65%) had a considerably higher proportion of iron supplement use compared with those in the lowest wealth category (11.42%). Access to healthcare services also played an important role, as women who reported that distance to a health facility was not a big problem (44.94%) had a higher proportion of iron supplement use than those who considered distance a major barrier ($p = 0.002$). Similarly, husband's employment status was strongly associated with iron supplement use ($p < 0.001$), with a substantially higher proportion of use reported among women whose husbands were employed (70.86%). Maternal health service utilization showed the strongest association with iron supplement use. Women who adequately utilized ANC services (39.03%) and those who partially utilized ANC services (45.07%) were more likely to use iron supplements compared with those who did not utilize ANC services (15.90%) ($p < 0.001$). In addition, media exposure was significantly associated with iron

supplement use ($p = 0.005$), with higher utilization observed among women who had exposure to media (44.59%) compared with those who had no exposure.

In contrast, age, place of residence, family size, parity, age at first birth, and respondent employment status were not significantly associated with iron supplement use ($p > 0.05$). These findings suggest that socioeconomic status, access to healthcare services, and exposure to health information play important roles in influencing iron supplement utilization among women in the study population.

Table 2: Bivariate Analysis of Iron Supplement Use. Predictors

Variable	Category	Non-use (%)	Iron supplement use (%)	χ^2 / F	p-value
Age	15–19	4.95	5.00	2.19	0.056
	20–24	17.02	15.28		
	25–29	26.98	32.70		
	30–34	21.63	25.72		
	35–39	18.49	14.42		
	40–44	8.28	5.82		
	45–49	2.64	1.05		
Region	Awdal	9.30	9.62	13.24	<0.001
	Marodijeh	22.23	35.82		
	Sahil	5.13	5.23		
	Togdheer	20.21	34.47		
	Sool	19.30	5.91		
	Sanaag	23.83	8.95		
Residence	Rural	56.26	58.73	0.24	0.627
	Urban	43.74	41.27		
Education level	No education	85.48	66.01	12.75	<0.001
	Primary	11.56	21.86		
	Secondary	1.78	7.79		
	Higher	1.18	4.34		
Family size	<4 members	21.60	19.32	0.79	0.377
	≥ 4 members	78.40	80.68		
Wealth index	Lowest	38.02	11.42	29.52	<0.001
	Second	16.65	8.67		
	Middle	10.50	10.11		
	Fourth	14.90	23.15		
	Highest	19.92	46.65		
Parity	Primipara	11.26	15.23	2.12	0.125
	Multipara	43.98	45.66		
	Grand multipara	44.76	39.10		
Age at first birth	<20	46.20	39.04	1.82	0.161
	20–24	36.49	41.10		
	25–30	14.98	17.30		
	≥ 30	2.33	2.56		

Distance to facility	Big problem	65.82	55.06	9.87	0.002
	Not a big problem	34.18	44.94		
Husband employment	Yes	37.40	70.86	50.90	<0.001
	No	61.73	28.58		
	Don't know	0.87	0.56		
Respondent employment	Yes	0.64	0.52	0.04	0.836
	No	99.36	99.48		
ANC visits	Not utilized	73.99	15.90	132.13	<0.001
	Partially utilized	17.97	45.07		
	Adequately utilized	8.03	39.03		
Media exposure	No	65.09	55.41	8.25	0.005
	Yes	34.91	44.59		

Machine learning algorithms performance

Table 3 presents the predictive performance of different machine learning algorithms used to classify iron supplement use. Overall, the models demonstrated moderate to high predictive performance across the evaluated metrics. Among the algorithms, the Support Vector Machine (SVM) achieved the highest accuracy (82.6%), followed by logistic regression (81.7%) and random forest (80.6%), while the decision tree model showed the lowest accuracy (75.6%). In terms of precision, logistic regression produced the highest value (66.5%), closely followed by SVM (66.0%) and random forest (65.7%), whereas the decision tree had the lowest precision (55.7%). Regarding recall (sensitivity), SVM demonstrated the best performance (74.5%), indicating a stronger ability to correctly identify women who used iron supplements, while K-nearest neighbors (45.2%) showed the lowest recall. For specificity, KNN achieved the highest value (90.2%), suggesting a better ability to correctly classify non-users of iron supplements, followed by XGBoost (88.6%) and random forest (88.2%). The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) values were relatively similar for most models, with logistic regression (85.3%), random forest (85.2%), and SVM (85.0%) demonstrating the highest discriminatory power. In comparison, the decision tree model had the lowest AUC (68.1%). Overall, these results indicate that SVM, logistic regression, and random forest provided the most reliable predictive performance for identifying iron supplement use among the study population.

Table 3: Predictive performance of machine learning algorithms

Metric	Logistic Regression	Random Forest	SVM	XGBoost	Naive Bayes	KNN	Decision Tree
Accuracy (%)	81.7	80.6	82.6	79.2	78.1	77.9	75.6
Precision (%)	66.5	65.7	66.0	64.2	59.5	63.4	55.7

Recall (%)	66.5	60.1	74.5	54.3	61.7	45.2	51.6
Specificity (%)	87.4	88.2	85.6	88.6	84.2	90.2	84.6
AUC-ROC (%)	85.3	85.2	85.0	83.7	81.8	80.3	68.1

Figure 2 presents the comparison of machine learning model performance across several evaluation metrics, including accuracy, precision, recall, and AUC-ROC. The results indicate that the Support Vector Machine (SVM) model achieved the highest overall performance, particularly in terms of accuracy (82.6%) and recall (74.5%), demonstrating a strong ability to correctly identify users of iron supplements. Logistic regression and random forest also showed strong performance, with high values across most evaluation metrics and AUC-ROC values exceeding 85%, indicating good discriminative ability. In contrast, the decision tree model exhibited the lowest performance across most metrics, particularly in accuracy and AUC-ROC, suggesting weaker predictive capability compared with the other models. Overall, the findings suggest that SVM, logistic regression, and random forest provided the most reliable and robust predictive performance for identifying iron supplement utilization in the study population.

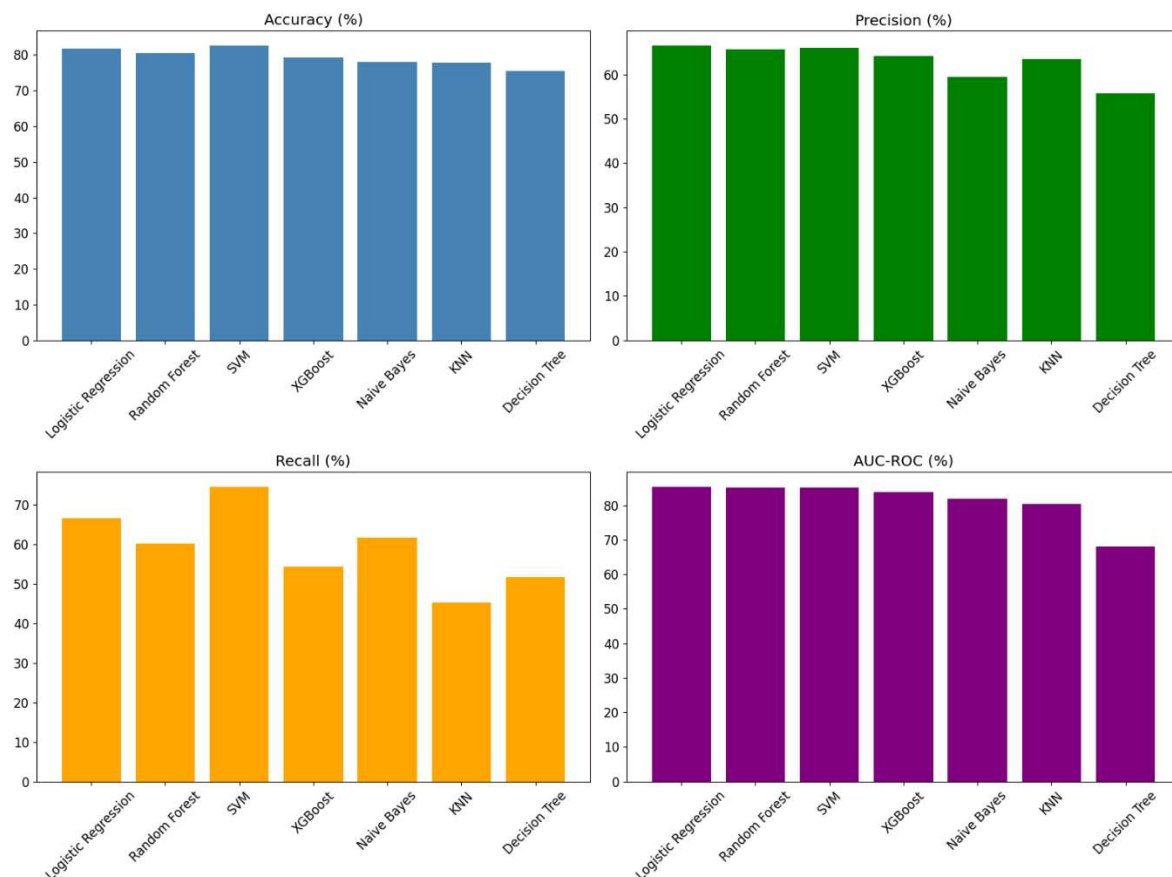


Figure 2 Comparison of machine learning model performance across evaluation metrics

Table 4 presents the confusion matrix results for the evaluated machine learning models used to classify iron supplement use. The results show variations in the ability of the models to correctly identify users and non-users of iron supplements. The Support Vector Machine (SVM) model achieved the highest number of true positives (TP = 288), indicating a stronger capability to correctly identify women who used iron supplements. Similarly, the Random Forest (RF) and XGBoost (XGB) models also demonstrated strong performance with 280 and 272 true positives, respectively. In contrast, Logistic Regression (LR) recorded the lowest number of true positives (217). Regarding false positives (FP), the logistic regression model produced the highest value (198), suggesting a greater tendency to incorrectly classify non-users as users. In comparison, SVM showed the lowest number of false positives (127), indicating better discrimination between users and non-users. For false negatives (FN), which represent women who used iron supplements but were incorrectly classified as non-users, SVM recorded the highest value (230), whereas Random Forest and XGBoost showed the lowest values (121). In terms of true negatives (TN), Random Forest and XGBoost achieved the highest counts (303), indicating a stronger ability to correctly identify women who did not use iron supplements, followed closely by KNN (302) and Decision Tree (290). Overall, the confusion matrix results suggest that SVM, Random Forest, and XGBoost demonstrated relatively strong classification performance, particularly in identifying true cases of iron supplement use and correctly distinguishing non-users. These findings highlight the comparative strengths of ensemble and kernel-based learning methods in predicting iron supplement utilization within the study population.

Table 4: Confusion matrix results for the evaluated machine learning models.

Metric	LR	SVM	KNN	DT	RF	GB	XGB
TP	217	288	240	272	280	247	272
FP	198	127	175	143	135	168	143
FN	170	230	122	134	121	161	121
TN	254	194	302	290	303	263	303

Figure 3 illustrates the Receiver Operating Characteristic (ROC) curve comparison of the machine learning models used to predict iron supplement use. The ROC curves demonstrate the trade-off between sensitivity (true positive rate) and specificity (false positive rate) across different classification thresholds. The results show that logistic regression achieved the highest discriminatory performance with an AUC of 0.853, followed closely by random forest (AUC = 0.852) and support vector machine (SVM) (AUC = 0.850), indicating strong predictive ability in distinguishing between iron supplement users and non-users. Similarly, XGBoost (AUC = 0.837) and Naive Bayes (AUC = 0.818) also demonstrated good classification performance, although slightly lower than the top-performing models. The K-nearest neighbors (KNN) model showed moderate performance with an AUC of 0.803, while the decision tree model exhibited the lowest predictive performance (AUC = 0.681) among the evaluated algorithms. Overall, the ROC analysis suggests that logistic regression, random forest, and SVM provided the most reliable classification performance, indicating their suitability for predicting iron supplement use in the study population.

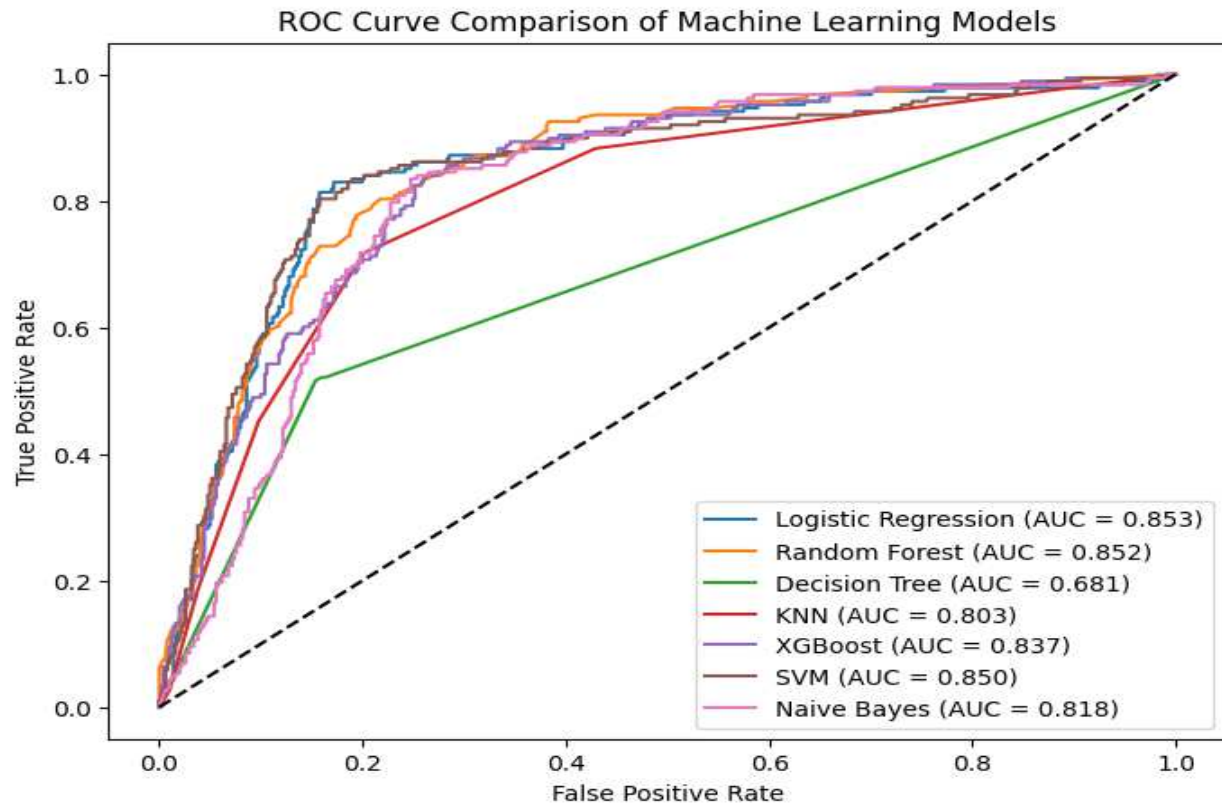


Figure 3: ROC curve comparison of machine learning models for predicting iron supplement use.

Feature Importance

Figure 4 shows the most important predictors of iron supplement utilization identified by the machine learning model. The results indicate that antenatal care (ANC) utilization, particularly non-utilization of ANC visits, was the most influential predictor of iron supplement use. Other important factors included husband's employment status, partial ANC utilization, urban residence, age at first birth below 20 years, media exposure, and perceived distance to a health facility. Additionally, parity, household size, and household wealth status were also identified as relevant predictors. Overall, the findings suggest that maternal healthcare utilization, socioeconomic factors, and access to health information play a significant role in influencing iron supplement use among women.

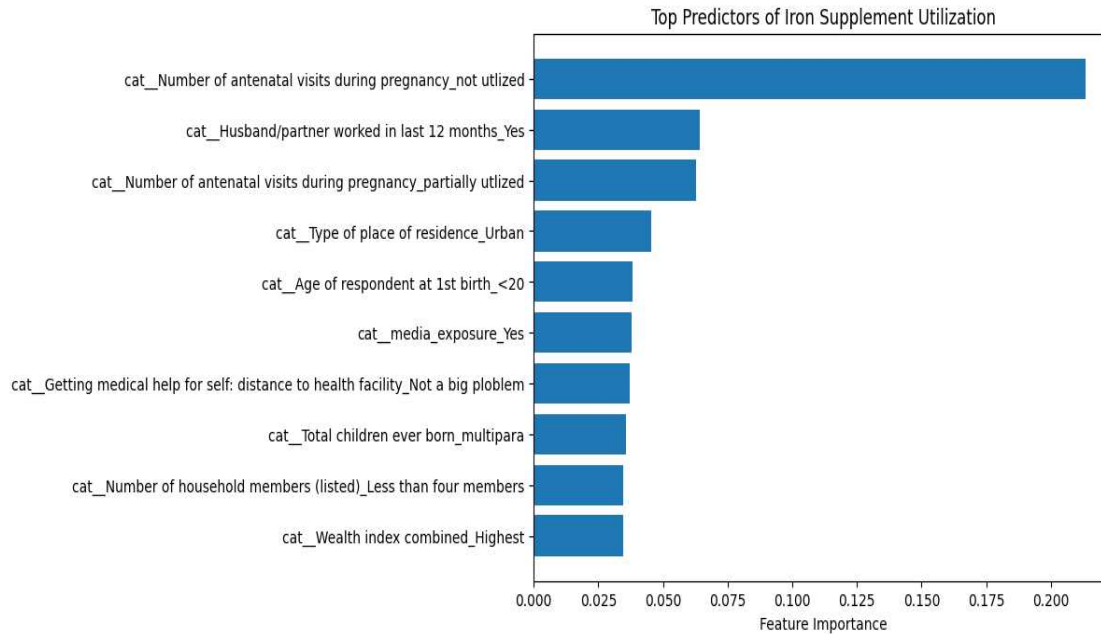


Figure 4: Feature importance of the top predictors of iron supplement utilization identified by the machine learning model.

Figure 5 compares the confusion matrices of the machine learning models used to predict iron supplement utilization. The results show that most models correctly classified a large proportion of non-users, with high true negative values across the models. The Support Vector Machine (SVM) and logistic regression models demonstrated relatively better balance between correctly identifying both users and non-users of iron supplements. In contrast, the decision tree and K-nearest neighbors (KNN) models showed comparatively lower performance in identifying true cases of iron supplement use. Overall, the confusion matrix comparison indicates that SVM, logistic regression, and random forest provided more reliable classification performance for predicting iron supplement utilization in the study population.

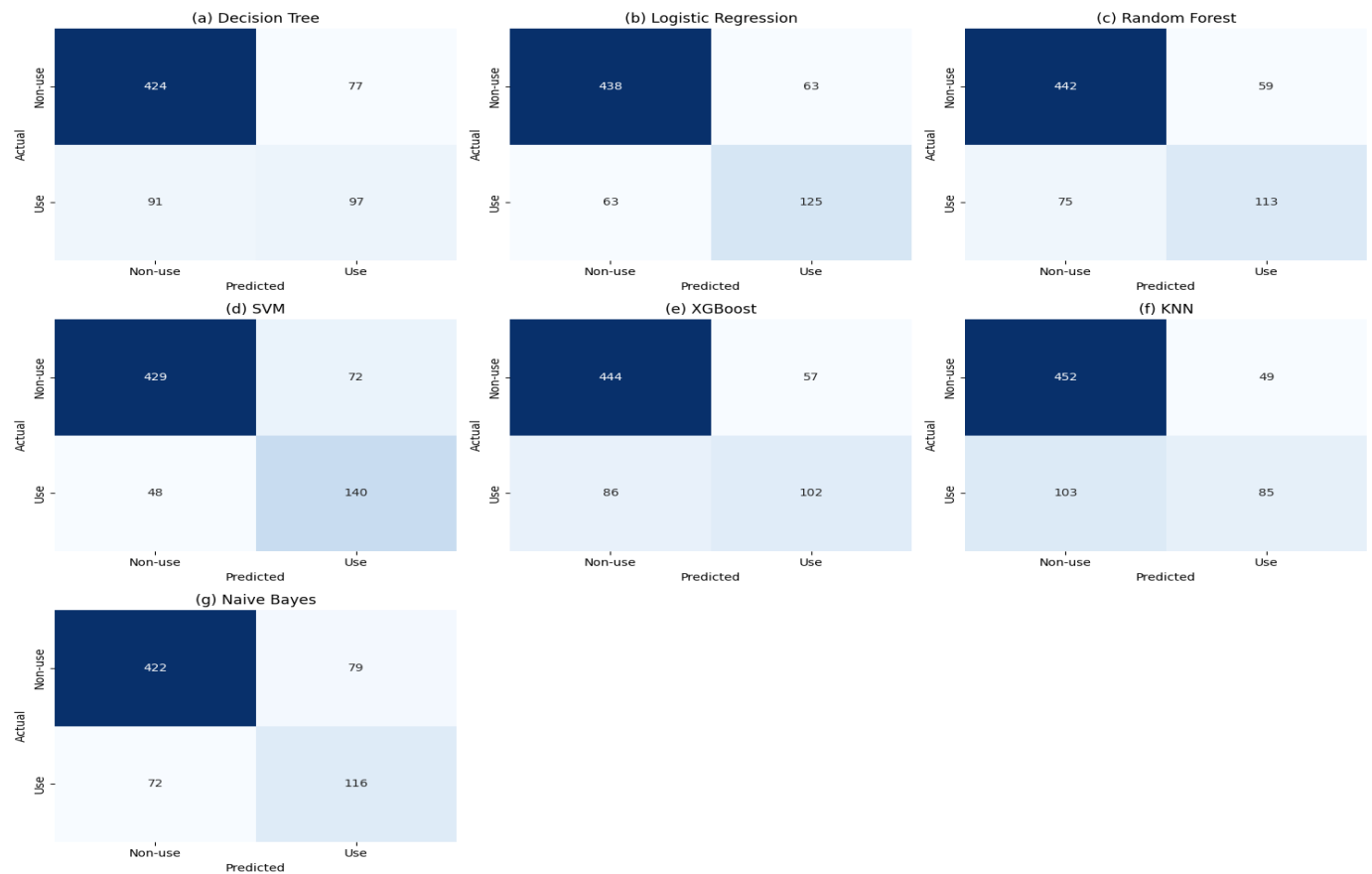


Figure 5: Confusion matrix comparison of machine learning models for predicting iron supplement utilization.

Figure 6 compares the importance of key predictors across different machine learning models, including logistic regression, random forest, decision tree, and support vector machine (SVM). The results consistently indicate that antenatal care (ANC) utilization was the most influential predictor of iron supplement use across all models. This was followed by media exposure and parity, which also demonstrated substantial contributions to the prediction of iron supplement utilization. Additionally, age at first birth and distance to a health facility showed moderate importance in influencing the model predictions. The consistency of these predictors across multiple algorithms suggests that maternal healthcare utilization, access to health information, and reproductive characteristics are key determinants of iron supplement use among women in the study population.

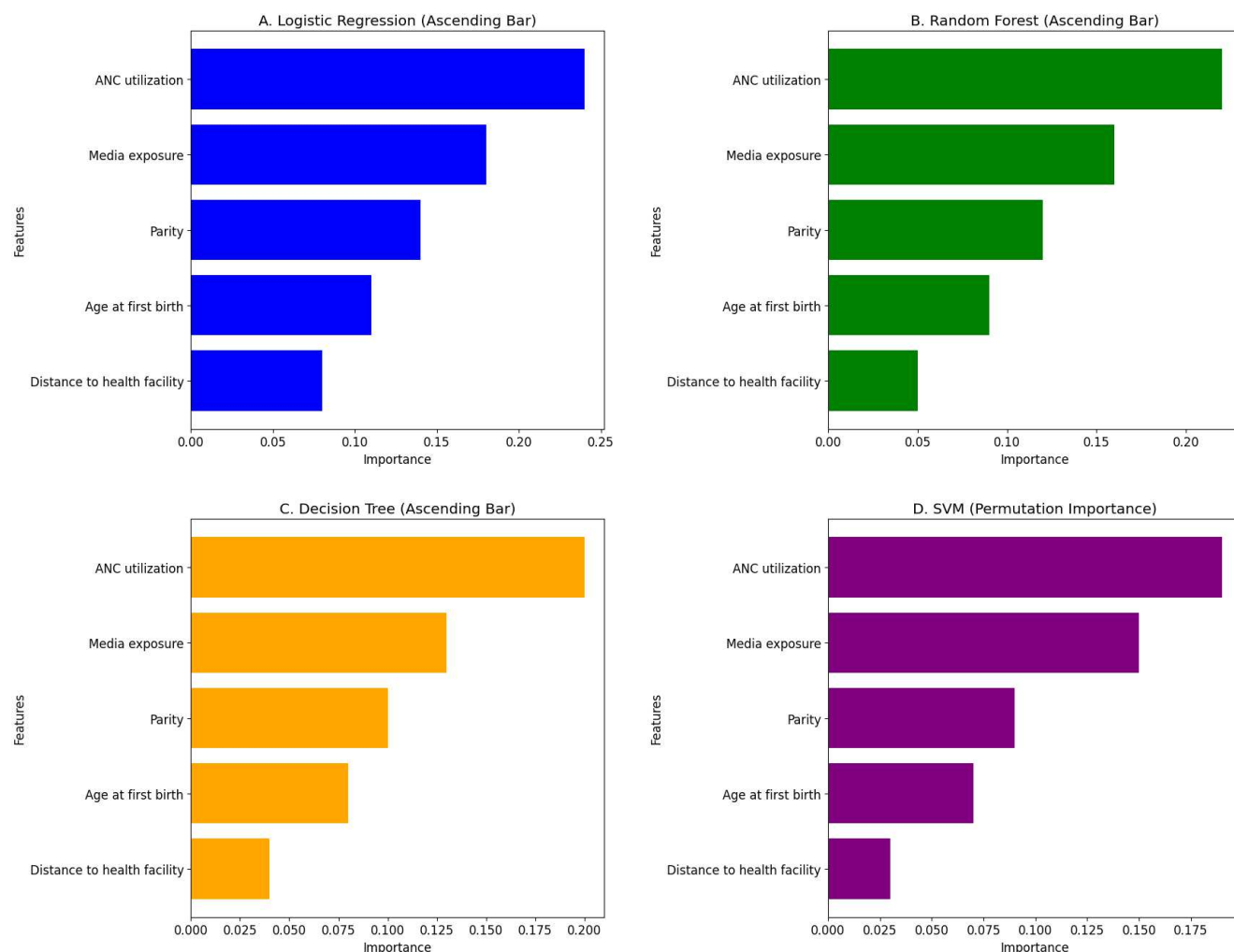


Figure 6: Feature importance comparison across machine learning models.

Discussion

This study applied a suite of supervised machine learning algorithms to examine the determinants of iron supplement utilization among pregnant women in Somaliland using the nationally representative 2020 Somaliland Demographic and Health Survey. By integrating traditional epidemiology analysis with advanced predictive modeling, the study not only identified the most influential predictors of iron supplements but also evaluated the comparative performance of multiple machine learning algorithms. The findings provide important insight into both the behavioral drivers of maternal micronutrient utilization and the methodological advantages of machine learning approaches in analyzing complex public health datasets. Situating these findings within the broader maternal health literature from low- and middle-income countries allows for a deeper understanding of the structural and behavioral factors influencing iron supplement uptake.

One of the most striking findings of this study is the low prevalence of iron supplement utilization, with only 27% of women reporting the use of iron supplementation during pregnancy. This level of utilization remains substantially below the global recommendations of the World

Health Organization, which emphasize routine iron and folic acid supplementation as a critical strategy for preventing maternal anemia and adverse birth outcomes. Similarly, low levels of iron supplementation have been reported in several sub-Saharan African countries(16-23), where structural Health system constraints, limited awareness of maternal nutrition, and barriers to accessing antenatal care services persist. The low coverage observed in this study, therefore, reflects broader systemic challenges within maternal health service delivery and highlights the need for strengthened public health strategies to improve micronutrient supplementation coverage. Research conducted by Peña-Rosas et al. demonstrated that preventive iron intake lowered the incidence of maternal anemia at delivery by 70% and decreased the likelihood of low birth weight by 16% in comparison to no supplementation(24). Furthermore, research by Habib et al. revealed that pregnant women who did not adhere to iron supplementation regimens faced a sixfold higher risk of developing anemia compared to those who were compliant(18). Significantly, the research reveals a wide disparity in iron supplement compliance among the 22 nations studied, with Senegal achieving the highest levels of adherence. These results align with earlier data from Seck and Jackson, who documented a 69% adherence rate within the Senegalese population(25).

Across the different machine learning models, antenatal care (ANC) utilization emerged as the most influential predictor of iron supplement use, consistently ranking as the top variable, as ANC visits serve as the primary delivery platform for essential maternal health interventions, including nutritional counseling and iron supplementation. Women who attend ANC visits are more likely to receive health education and direct access to micronutrient supplements, thereby increasing adherence to recommended maternal nutrition practices. The dominance of ANC in shaping health behaviors suggests that improving ANC coverage may substantially increase iron supplementation uptake (26, 27).socioeconomic and informational factors also played a meaningful role in predicting iron supplement utilization. In particular, media exposure and household socioeconomic status emerged as important determinants across several models. Media exposure likely enhances women's awareness of maternal nutrition and the benefits of micronutrient supplementation, thereby influencing health-seeking behavior. Similarly, women from relatively better socioeconomic backgrounds may have improved access to healthcare services and greater autonomy in making health-related decisions. Additionally, distance to health facilities appeared as a relevant predictor, highlighting the persistent influence of geographic accessibility on maternal health service utilization. These findings collectively reinforce the importance of addressing structural barriers to healthcare access while simultaneously improving public health communication strategies.(18, 20, 28-30).

Another important methodological insight from this study is the comparative performance of the machine learning algorithms. Among the evaluated models, support vector machine (SVM), logistic regression, and random forest demonstrated the strongest predictive performance, with high levels of accuracy and discriminative ability. The strong performance of both traditional statistical models and advanced machine learning algorithms suggests that iron supplement utilization is influenced by a combination of linear and non-linear relationships among predictors. Importantly, machine learning models provide additional analytical advantages by capturing

complex interactions between demographic, socioeconomic, and healthcare access variables that may not be fully captured using conventional regression approaches. This highlights the growing potential of machine learning techniques in advancing public health research and improving predictive modeling in maternal health studies.(31)

Finally, the findings of this study underscore the critical role of health system accessibility, maternal healthcare utilization, and information exposure in shaping iron supplement use among women. In contexts such as Somaliland, where healthcare infrastructure and service accessibility remain uneven across regions, maternal health behaviors are often determined more by structural constraints than by individual characteristics alone. Consequently, interventions aimed at improving iron supplementation coverage should prioritize expanding access to antenatal care services, strengthening community-based health education, and reducing geographic barriers to healthcare access. Addressing these systemic challenges may significantly enhance maternal nutrition and ultimately contribute to improved maternal and neonatal health outcomes.

Strengths and Limitations

The primary strength of this study lies in its novel integration of machine learning techniques with a nationally representative dataset (SLDHS 2020) to identify the key predictors of iron supplement utilization among women in Somaliland. By combining traditional epidemiological analysis with multiple supervised machine learning algorithms, the study provides a data-driven ranking of influential factors that can inform evidence-based maternal health policies. This analytical approach enables the detection of complex and potentially non-linear relationships between sociodemographic, reproductive, and healthcare access variables that may not be fully captured by conventional regression models.

However, several limitations should be acknowledged. First, the cross-sectional design of the dataset prevents the establishment of causal relationships between the identified predictors and iron supplement utilization. Second, the reliance on self-reported information may introduce recall bias, particularly for variables related to maternal healthcare utilization. Third, the analysis was restricted to the variables available in the SLDHS dataset, which limited the ability to examine additional contextual factors such as health system supply constraints, service quality, and individual knowledge regarding maternal nutrition. Finally, while machine learning models offer strong predictive capabilities, the interpretability of some algorithms remains limited, as their internal decision-making processes may function as a “black box.”

Conclusion

This study provides important evidence on the determinants and predictive patterns of iron supplement utilization among women in Somaliland using nationally representative data and machine learning approaches. The findings revealed that iron supplement use remains substantially low, indicating significant gaps in maternal micronutrient supplementation within the study population. Among the identified determinants, antenatal care utilization emerged as the most influential predictor, highlighting the central role of maternal healthcare engagement in promoting iron supplementation during pregnancy.

The machine learning analysis further demonstrated that support vector machine, logistic regression, and random forest models provided the strongest predictive performance, underscoring the value of integrating advanced analytical methods into public health research to better understand maternal health behaviors. The study also highlights the importance of health information exposure, socioeconomic conditions, and geographic accessibility to health services in shaping iron supplement utilization.

Overall, improving iron supplementation coverage in Somaliland will require strengthening antenatal care services, expanding maternal health education through mass media, and addressing structural barriers related to healthcare accessibility. Targeted public health interventions that focus on these areas may significantly improve maternal nutrition and contribute to better maternal and neonatal health outcomes.

List of Abbreviations

ANC (Antenatal Care); AUC (Area Under the Curve); AUC-ROC (Area Under the Receiver Operating Characteristic Curve); DHS (Demographic and Health Survey); FN (False Negative); FP (False Positive); IFAS (Iron and Folic Acid Supplementation); KNN (K-Nearest Neighbors); LR (Logistic Regression); ML (Machine Learning); RF (Random Forest); ROC (Receiver Operating Characteristic).

Declarations

Ethics approval and consent to participate

Ethical approval was not required because this study used publicly available, anonymized secondary data from the Demographic and Health Survey (DHS). I obtained permission to download the DHS dataset from the Central Statistical Agency (CSA) through a request at <https://microdata.nbs.gov.so/index.php/catalog/50>

Consent for publication

Not applicable

Data Availability

The Somaliland Demographic and Health Survey (SLHDS) 2020 datasets used in this study are publicly accessible and can be obtained through an online request at <https://microdata.nbs.gov.so/index.php/catalog/50>, specifying the purpose of the research. The datasets analyzed during the current study are also available from the corresponding author upon reasonable request.

Competing interests

The authors declare no competing interests.

Funding

This research received no external funding

Authors contributions

Author 1: conducted data cleaning, analysis, and interpretation.

Author 2: consulted on data analysis and publications.

Author 3: prepared the background section.

Author 4: wrote the discussion section.

Author 5: wrote the conclusion section

All authors reviewed and approved the final manuscript.

Acknowledgements

The authors would like to express their sincere gratitude to the Somaliland Ministry of Health Development and the Somaliland Demographic and Health Survey (SLDHS) 2020 team for granting access to the dataset used in this study. We also acknowledge to colleagues and reviewers who provided valuable comments and constructive feedback that helped improve the quality of this manuscript.

Reference

1. Kamau MW, Mirie W, Kimani S. Compliance with Iron and folic acid supplementation (IFAS) and associated factors among pregnant women: results from a cross-sectional study in Kiambu County, Kenya. *BMC public health*. 2018;18(1):580.
2. Fite MB, Roba KT, Oljira L, Tura AK, Yadeta TA. Compliance with Iron and Folic Acid Supplementation (IFAS) and associated factors among pregnant women in Sub-Saharan Africa: A systematic review and meta-analysis. *PLoS One*. 2021;16(4):e0249789.
3. Gebremedhin S, Samuel A, Mamo G, Moges T, Assefa T. Coverage, compliance and factors associated with utilization of iron supplementation during pregnancy in eight rural districts of Ethiopia: a cross-sectional study. *BMC public Health*. 2014;14(1):607.
4. 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):481.
5. Beressa G, Lencha B, Bosha T, Egata G. Utilization and compliance with iron supplementation and predictors among pregnant women in Southeast Ethiopia. *Scientific Reports*. 2022;12(1):16253.
6. Yousuf HJ, Egeh MH, Farah AA, Hussein AA, Omer AA. Determinants of Tetanus Toxoid Immunization Among Pregnant Women in Somaliland: Evidence from the 2020 Nationwide Survey Using a Zero-Inflated Negative Binomial Model. 2026.

7. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.
8. Mazzocco T, Hussain A. Novel logistic regression models to aid the diagnosis of dementia. *Expert Systems with Applications*. 2012;39(3):3356-61.
9. Breiman L, Friedman J, Olshen RA, Stone CJ. Classification and regression trees: Chapman and Hall/CRC; 2017.
10. Biau G, Scornet E. A random forest guided tour. *Test*. 2016;25(2):197-227.
11. Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995;20(3):273-97.
12. Abbafati C, MacHado DB, Cislaghi B, Salman OM, Karanikolos M, McKee M, et al. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. 2020.
13. Kombo OH, Kumaran S, Sheikh YH, Bovim A, Jayavel K. Long-term groundwater level prediction model based on hybrid KNN-RF technique. *Hydrology*. 2020;7(3):59.
14. Chen F-C, Jahanshahi MR. NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion. *IEEE Transactions on Industrial Electronics*. 2017;65(5):4392-400.
15. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction: Springer; 2009.
16. Kassa ZY, Awraris T, Daba AK, Tenaw Z. Compliance with iron folic acid and associated factors among pregnant women through pill count in Hawassa city, South Ethiopia: a community based cross-sectional study. *Reproductive health*. 2019;16(1):14.
17. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *American journal of epidemiology*. 2005;162(3):199-200.
18. Habib F, Habib Zein Alabdin E, Alenazy M, Nooh R. Compliance to iron supplementation during pregnancy. *Journal of Obstetrics and Gynaecology*. 2009;29(6):487-92.
19. Tegegne M. Compliance to prenatal Iron and Folic acid supplement and associated factors among women during pregnancy in South East Ethiopia: A Cross-sectional study. *Global Journal of Medical Research*. 2017;7(2).
20. Agegnehu G, Atenafu A, Dagne H, Dagne 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. *International journal of reproductive medicine*. 2019;2019(1):5863737.
21. Arega Sadore A, Abebe Gebretsadik L, Aman Hussen M. Compliance with iron-folate supplement and associated factors among antenatal care attendant mothers in Misha District, South Ethiopia: Community based cross-sectional study. *Journal of environmental and public health*. 2015;2015(1):781973.
 22. 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 research notes*. 2017;10(1):510.
 23. Sununtnasuk C, D'Agostino A, Fiedler JL. Iron+ folic acid distribution and consumption through antenatal care: identifying barriers across countries. *Public Health Nutrition*. 2016;19(4):732-42.
 24. Viteri F. Daily oral iron supplementation during pregnancy. *Reviews*. 1996.
 25. Seck BC, Jackson RT. Determinants of compliance with iron supplementation among pregnant women in Senegal. *Public health nutrition*. 2008;11(6):596-605.
 26. Getachew M, Abay M, Zelalem H, Gebremedhin T, Grum T, Bayray A. Magnitude and factors associated with adherence to Iron-folic acid supplementation among pregnant women in Eritrean refugee camps, northern Ethiopia. *BMC pregnancy and childbirth*. 2018;18(1):83.
 27. Tarekegn M, Wubshet M, Atenafu A, Derso T, Woretaw A. Antenatal care and mothers' education improved iron-folic acid adherence at Denbiya district health centers, Northwest Ethiopia: using pills count method. *Archives of Public Health*. 2019;77(1):30.
 28. Titilayo A, Palamuleni M, 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 Medical Journal*. 2016;28(1):1.
 29. Nisar YB, 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 and childbirth*. 2014;14(1):305.

30. Taye B, Abeje G, Mekonen A. Factors associated with compliance of prenatal iron folate supplementation among women in Mecha district, Western Amhara: a cross-sectional study. *Pan African Medical Journal*. 2015;20(1).
31. Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of statistics*. 2001:1189-232.