Discover Public Health

Research

Socioeconomic and demographic factors associated with anaemia among women of reproductive age in Zimbabwe: a supervised machine learning approach

Garikayi Chemhaka[1] · Elliot Mbunge[2] · Tafadzwa Dzinamarira[3] · Godfrey Musuka[4] · John Batani[5] · Benhildah Muchemwa[6] · Stephen Fashoto[6] · Munyaradzi Mapingure[7] · Rutendo Birri Makota[8] · Ester Petrus[9]

Received: 3 July 2024 / Accepted: 25 March 2025
Published online: 07 April 2025

Abstract
Anaemia affects approximately one-third of women of reproductive age globally, with the highest burden observed in resource-limited countries. Therefore, this study aimed to determine the socioeconomic and demographic factors associated with anaemia and predict anaemia among women in Zimbabwe. Using nationally representative, cross-sectional data from the 2015 Zimbabwe Demographic and Health Survey (DHS), a dataset from a sample of 5412 women of reproductive age was analyzed. The Chi-square test and multivariate logistic regression were employed to identify independent predictors of anaemia, while Elastic Net was used for feature importance scoring. To address the class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied. The prevalence of anaemia among women in Zimbabwe was 24.1%. Multivariate logistic regression revealed significant associations between anaemia and several factors, including older age (35–49 years) (adjusted Odds Ratio [aOR] = 1.31), marital status (being married) (aOR = 0.72), higher education (aOR = 0.47), middle household wealth (aOR = 1.32), professional occupation (aOR = 1.60), current use of modern contraceptives (aOR = 0.59), and overweight/obesity (aOR = 0.56). The highest burden was observed in Matabeleland South province (aOR = 3.44). Among prediction models, the random forest classifier outperformed K-Nearest Neighbors (KNN) and decision trees, achieving an accuracy of 74%, recall of 78%, F1-score of 75%, precision of 72%, and an Area Under the Curve (AUC) of 81.5%. Targeted interventions focusing on key socioeconomic and demographic characteristics could help reduce anaemia in women of reproductive age. Predictive models can aid healthcare practitioners in identifying at-risk individuals and implementing timely interventions to mitigate the impact of anaemia.

Keywords Anaemia · Machine learning · Survey · Socioeconomic · Demographic · Logistic regression · Random Forest · K-Nearest Neighbors · Decision Trees

✉ Elliot Mbunge, mbungeelliot@gmail.com; Garikayi Chemhaka, bgchem@gmail.com; Tafadzwa Dzinamarira, anthonydzina@gmail.com; Godfrey Musuka, gmusuka@3ieimpact.org; John Batani, jonj188@gmail.com; Benhildah Muchemwa, benhildahmuchemwa@gmail.com; Munyaradzi Mapingure, pmapingure@yahoo.co.uk; Rutendo Birri Makota, rutendobbirri@gmail.com; Ester Petrus, esterpetrus@gmail.com | [1]Department of Statistics and Demography, Faculty of Social Sciences, University of Eswatini, Private Bag 4, Kwaluseni Campus, Kwaluseni, Eswatini. [2]Division of Research, Innovation and Engagement, Mangosuthu University of Technology, Umlazi, Durban 4031, South Africa. [3]School of Health Systems and Public Health, University of Pretoria, Pretoria 0002, South Africa. [4]International Initiative for Impact Evaluation, Harare, Zimbabwe. [5]Faculty of Engineering and Technology, Botho University, Maseru, Lesotho. [6]Department of Computer Science, Faculty of Science and Engineering, University of Eswatini, Private Bag 4, Kwaluseni Campus, Kwaluseni, Eswatini. [7]ICAP in Zimbabwe, Harare, Zimbabwe. [8]Department of Biological Sciences and Ecology, University of Zimbabwe, Harare, Zimbabwe. [9]Software Department, Faculty of ICT, International University of Management, 21-31 Hercules Street, Dorado Park Ext 1, Windhoek, Namibia.

Abbreviations

BMI	Body Mass Index
SDGs	Sustainable Development Goals
ZDHS	Zimbabwe Demographic and Health Survey
ZIMSTAT	Zimbabwe National Statistics Agency
AOR	Adjusted odds ratio
CI	Confidence intervals
ML	Machine learning

1 Introduction

Anaemia is a serious challenge for the global population and public health, worst in resource-poor nations [1–3], and in particular for children under 5 years and women of reproductive age 15–49 years [4–6]. Globally, one-third of women of reproductive age are estimated to be anaemic [4, 6]. Women in their reproductive span are vulnerable since they require more iron when pregnant, breastfeeding, and menstruating [4, 6, 7]. Anaemia prevalence is classified into three categories: mild, moderate, and severe, with values ranging from 5–19.9%, 20–39.9%, and 40% or more, respectively [8, 9]. Moderate-to-severe anaemia is common in low and middle-income countries [3, 6, 7]. Anaemia is a condition characterised by low haemoglobin concentration due to reduced red blood cell levels in tissues due to low oxygen intake [10]. Haemoglobin is an iron-rich protein carrying oxygen. When haemoglobin levels are lowered below normal levels, it impairs tissues and causes fatigue [5, 6]. For non-pregnant women, the haemoglobin threshold levels for anaemia are 11.0–11.9 (mild), 8.0–10.9 (moderate), and < 8.0 (severe) grams per deciliter (g/dl), while for pregnant women, they are 10.0–10.9, 7.0–9.9, and < 7.0 g/dl, respectively [6, 11, 12].

Anaemia is caused by malaria and parasitic infections, haemoglobinopathies (caused by genetic defects), and iron deficiency [13–15]. Maternal anaemia is associated with adverse child and maternal health outcomes including low birth weight, stillbirth, infant mortality, preeclampsia, postpartum haemorrhage, depression, and maternal mortality [5, 16]. The risk for anaemia is elevated for women in resource-poor nations because of nutritional deficiencies (e.g. folate, vitamin B12, lack of iron) and clinical or parasitic infectious diseases (e.g. HIV/AIDS, tuberculosis, malaria, hookworm infection) compared to those in high-income countries [2, 16]. In addition to anaemia having adverse nutritional and disease outcomes resulting in maternal morbidities and ultimately death, it also has demographic, socioeconomic, and other health implications. Previous empirical studies have demonstrated the association between maternal anaemia and several distal factors including age [1, 7, 13], age at first birth [17], parity [18–20], marital status [4, 16], religion [6, 20], education, household wealth, occupation [7, 21–24]. Other factors associated with maternal anaemia are access to media, residence [4, 21, 25, 26], type of toilet facility, source of drinking water, type of cooking fuel [7, 22, 27], modern contraceptive use, body mass index, smoking, pregnant status, breastfeeding status, pregnancy loss, slept und, and malaria prevention using a mosquito net [1, 7, 22, 23, 28].

Zimbabwe is one of the countries in the southern Africa region with a notable prevalence of anaemia estimated at 27% in 2015 dropping from 38% in 2006 [29]. The present progress towards Sustainable Development Goal (SDG) 2.2.3 of the United Nations, which calls for reducing the prevalence of anaemia in women of reproductive age by half (50%) by 2025 [2, 7], is not adequate. As indicated, a prevalence of anaemia above 5% is of public health significance. Therefore, the general moderate maternal anaemia public health problem in Zimbabwe raises concerns about the causes of these levels on a national scale. However, we lack detailed knowledge of the associated factors of maternal anaemia in the population. A recent study in Zimbabwe analysed limited variables and found HIV, residence, breastfeeding, and pregnant women were associated with anaemia in women of reproductive age [25]. Other studies have focused on Zimbabwe in multicountry analysis by using pooled data. Some of these studies have looked at subgroups of women of reproductive age such as those pregnant [30, 31] or lactating [32]. To effectively target public health interventions, our study sought to estimate the prevalence of anaemia and identify risk factors associated with anaemia among women of reproductive age utilising a wide range of individual-level socioeconomic and demographic characteristics mentioned in the literature.

2 Materials and methods

2.1 Study setting

Zimbabwe is a lower-middle-income, land-linked country situated in the southern part of Africa [32]. The country has eminent emigration and economic challenges characterised by hyperinflation, poor healthcare, and high unemployment [33]. According to the 2015 Zimbabwe Demographic and Health Survey report [29], the population of Zimbabwe is relatively young with a fertility rate of 4.0 children per woman, and an average household size of 4.1 members in 2015. Further, about 3 in 5 people (men and women) reside in rural areas, and 3 in 5 households in rural households belong to the poorest and poor wealth status. The burden of HIV remains a public health threat with prevalence much higher among women (16.7%) than men (10.5%) aged 15–49 years [29]. The country is divided into 10 regions or provinces: Manicaland, Mashonaland Central, Mashonaland West, Mashonaland East, Midlands, Matabeleland South, Matabeleland North, Masvingo, Bulawayo, and Harare. The latter is the capital city. Masvingo, Bulawayo, and Matabeleland provinces are situated in drought-prone regions 4 and 5 [25, 34].

2.2 Data source, sampling technique, and population

A cross-sectional design was used with a nationally representative woman of reproductive age (15–49) from the recent Zimbabwe Demographic and Health Survey (ZDHS) conducted in 2015. The data was collected using a two-stage sampling method. In the first stage, the primary sampling units were drawn with a probability proportional to the size within each stratum. In the second stage, households were selected using a systematic sampling method in the selected primary sampling unit. In each household, all women aged 15–49 years were included in the study [29]. For this study, from a total of 6132 women in the dataset, a sample of 5412 weighted complete cases for women of reproductive age was used for women who had anaemia. Therefore, missing cases and nonresponses from the selected outcome variable in the study were excluded from the analysis.

2.3 Measures

2.3.1 Outcome

The outcome measure for this study was anaemia, irrespective of its cause. A haemoglobin threshold level of below 11 g/dl for pregnant women and 12 g/dl for non-pregnant women was the indicator of anaemia [6, 29]. For this study, mild, moderate, and severe anaemia in the DHS dataset (V457) were classified as anaemic (code "1") and otherwise as non-anaemic (code "0").

2.3.2 Explanatory variables

We selected 21 explanatory variables from existing literature which include age (15–24, 25–34, 35–49), age at first birth (< 18, 18+), marital status (not in union, married), parity (1–2, 3–5, 6+), religion (none/other, Catholic, Protestant, Pentecostal, Apostolic), education (primary/less, secondary, higher), household wealth (poor, middle, rich), occupation (unemployed, agricultural/manual labour, sales/services, professional/technical/clerical, other), modern contraceptive use (no, yes), access to media (no, yes), type of toilet facility (improved, unimproved), source of drinking water (improved, unimproved), type of cooking fuel (low-polluting, high-polluting), body mass index (< 18.5 kg/m^2: underweight, 18.5–24.99 kg/m^2: normal, ≥ 25 kg/m^2: overweight and obese), smoking (no, yes), pregnant (no, yes), breastfeeding (no, yes), slept under a mosquito net (no, yes), ever terminated pregnancy (no, yes), area (urban, rural) and province/region (Manicaland, Mashonaland Central, Mashonaland West, Mashonaland East, Midlands, Matabeleland South, Matabeleland North, Masvingo, Bulawayo, and Harare). The household wealth variable is found in the DHS dataset classified as five quintiles rankings (V190) and is derived using principal component analysis (PCA) from goods and items assets in the household [35]. The wealth quintiles were re-categorised as poor (poorest, poor), middle, and rich (richer and richest). Sources of drinking water (V113) from piped, tap, borehole, protected spring or well, and bottled water were categorised as improved and otherwise as unimproved. The type of toilet facility (V116) used such as flush, pit latrine with a slab, and ventilated improved pit latrine were categorised as improved and otherwise unimproved. The type of cooking fuel (V161) was classified as low-polluting from electricity, gas, kerosene, and no cooking at home and otherwise categorised as high-polluting.

2.4 Statistical analysis

The study population was explained using summary statistics (frequencies and percentages) and cross-tabulations. The chi-square test was used to test the association between each explanatory variable and anaemia. Regardless of their significance, all explanatory variables were included in a multivariate logistic regression model to identify independent predictors of anaemia among women aged 15–49 years. A parsimonious logistic regression model with all statistically significant explanatory variables was set as the final model using stepwise regression modelling. The results were presented as adjusted odds ratios (aOR) with 95% Confidence Intervals (CI). There was no evidence of collinearity after a multicollinearity test utilising the Variance Inflation Factor (VIF) since VIF < 10 (VIF mean = 2.63, VIF Min = 1.16, Max = 7.53). The accuracy of the model was examined using the Hosmer–Lemeshow test (P > 0.05). The resultant or final model was adequate and of good fit (P = 0.4968). Further, the model accuracy was confirmed by a statistically significant Wald Chi-square test value of 156.2 (P < 0.001). Results from the multivariate logistic regression and the Chi-square test were both determined to be significant at a P-value of < 0·05. Stata's 'svyset' command was used for weighting and stratifying the data to account for the complex survey design of the DHS.

2.5 Machine learning approaches for predicting anaemia

After identifying significant variables using Chi-square and multivariate logistic regression model, the study further applied advanced machine learning (ML) algorithms to predict anaemia among women of reproductive age. Unlike statistical models, machine learning models have the predictive capacity to determine the likelihood of a woman having anaemia using significant determinants/variables identified through statistical methods shown in Tables 2 and 3. In addition, each significant variable is given the importance score using Elastic Net, as shown in Fig. 3. This provides experts in public health to understand the degree of significance associated with each variable. Machine learning (ML) is a subfield of artificial intelligence that can learn and extract interesting patterns from data without explicitly programmed [36–38]. The application of machine learning algorithms in public health has made significant progress [39–41], especially in modelling risk factors associated with anaemia and predicting anaemia. For instance, Meitei et al. [42] applied ML algorithms on National Family Health Survey (NFHS-4) data to predict child anaemia with better accuracy. Vohra et al. [43] also applied ML algorithms such as decision trees, logistic regression, Naïve Bayes, random forest, support vector machine (SVM), and multilayer perceptron to diagnose anaemia. In India and Bangladesh, logistic regression was used to identify risk factors associated with women's anaemia in limited resource areas by Jana et al. [44]. Appiahene et al. [45] also applied convolutional neural networks, KNN, DT, Naïve Bayes, and SVM to detect iron deficiency anaemia and achieve high accuracy. In Ethiopia, Dejene et al. [46] applied decision trees, random forest, CatBoost, and extreme gradient boosting to predict the level of anaemia among pregnant women, with CatBoost achieving an accuracy of 97.6%. Therefore, in this study, we utilised important features associated with anaemia among women of reproductive age selected using a parsimonious multivariate logistic regression model. The selected features include age, marital status, education, household wealth, occupation, modern contraceptive use, body mass index, and province/region (shown in Table 3). All missing data were deleted, and the categorical data were converted into numerical data using the data label encoding technique. The target class had an uneven distribution of samples (Anaemia-No: 4075 and Yes: 1337) as shown in Fig. 1. An imbalanced data in the target class can cause bias in machine learning models [47].

To address the class imbalance problem, the study applied the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is a popular oversampling technique that uses random replication of minority class instances to balance class distribution [48]. SMOTE uses the KNN technique to generate synthetic data [49]. The process begins by selecting random data points from the minority class. SMOTE further applies KNN to identify the nearest data point and creates synthetic data by interpolating between the selected point and its nearest neighbour [50]. This process continues until the minority class is balanced with the majority class in the dataset. Pearson correlation matrix was used to determine multicollinearity and understand the relationships between the selected explanatory variables, as shown in Fig. 2.

Furthermore, the study applied Elastic Net to determine feature importance scores, as shown in Fig. 3. Elastic Net is a regularisation technique commonly used in linear regression models to handle multicollinearity and perform feature selection [51]. Elastic net combines both Lasso regularisation (L1) and Ridge (L2) regularisation methods, and leverages on

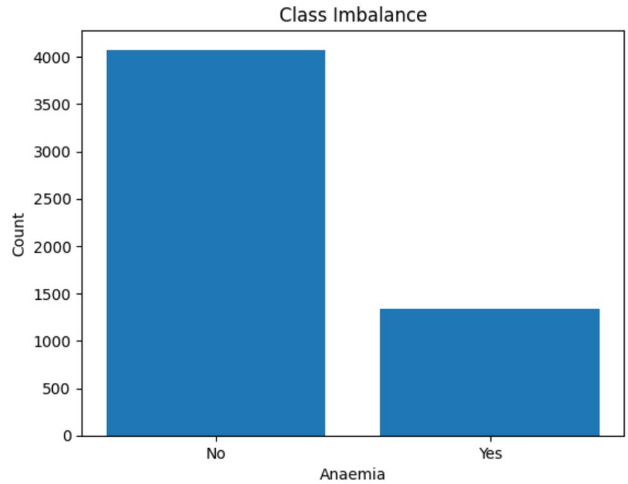

Fig. 1 Class imbalance

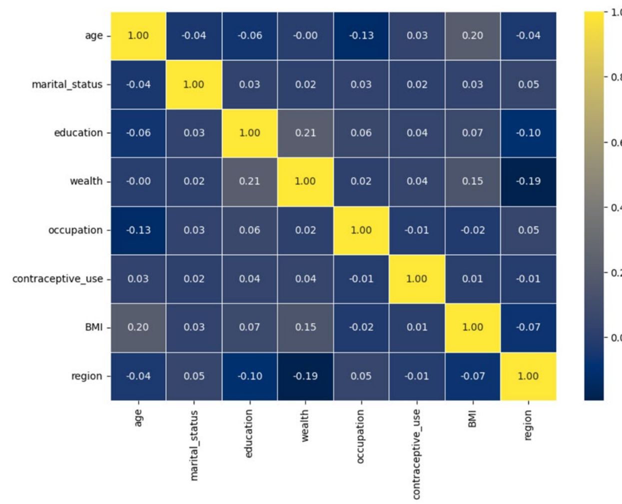

Fig. 2 Pearson correlation heatmap

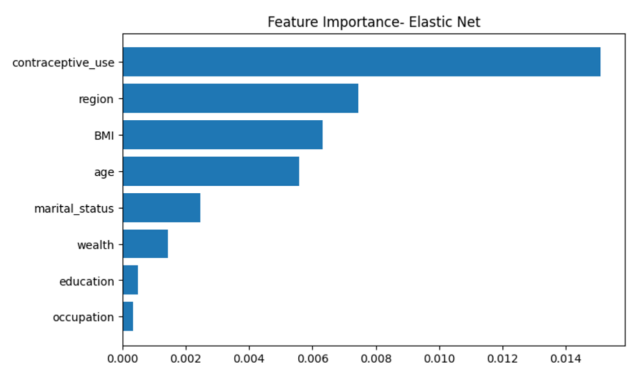

Fig. 3 Feature importance score

L1 for feature selection [52]. The feature importance score, based on Elastic Net indicates that contraceptive use, region, BMI, and age are the most important features in predicting anaemia among women of reproductive age.

After determining the feature importance score, the study split the dataset into training and testing sets using the ratio of 80% and 20%, respectively. The study further applied K-Nearest Neighbors (KNN), Decision Trees (DT) and Random Forest (RF) to predict anaemia among women of reproductive age. The K-Nearest Neighbors classifier is a supervised machine-learning algorithm that can used to solve classification problems. It is a non-parametric algorithm that classifies new data points based on their similarity to the training data [53]. A decision tree classifier is a popular supervised machine learning algorithm that creates a tree-like model of decisions and their possible consequences based on the features of the input data [46]. It builds a decision tree based on the provided training data, where each internal node represents a feature or attribute, each branch represents a possible value of that feature, and each leaf node represents a decision [54]. A random forest classifier is an ML algorithm that combines multiple decision trees to make predictions. It uses decision trees, where each tree is trained on a random subset of the training data and features [46]. Random Forest uses bootstrap aggregating (or bagging) to construct trees during the training process. This ensures that each tree has a slightly different perspective on the data [55].

3 Results

3.1 Description of the participants

The descriptive analysis in Table 1 shows that slightly above half (50.8%) of women were aged 25–34. The majority of women reported that they were currently married (85.4%), had a first birth aged 18 or older (73.9%), had completed secondary education (62.6%), and resided in rural areas (68.6%). Nearly a quarter (24.1%) of the women were anaemic (Table 1). Table 2 results reveal that marital status, modern contraceptive use, body mass index, pregnant status, and the province/region of residence were independently associated with anaemia. The prevalence of maternal anaemia was highest among women who smoke (46%). Notably, anaemia was high in the province of Matabeleland South (41%), among pregnant women (33.4%), among not-in-union/non-married women (33.7%), women not using modern contraceptives (32.3%), among underweight women (31.3%) and among Catholics (30.6%). Anaemia was lowest in the province of Manicaland (16.2%) (Table 1).

3.2 Exploratory results using chi-square

See Table 2.

3.3 Multivariate analyses with logistic regression

The results with adjusted odds ratios (aOR) and their associated 95% confidence intervals (95% CI) indicate that age, marital status, education, household wealth, occupation, modern contraceptive use, body mass index, and province/region were associated with anaemia. Although pregnant status was statistically significant in the bivariate analysis (Table 2), it was not so after accounting for the influence of other variables. The rest of the selected independent variables were not associated with anaemia (see Tables 1 and 2). After controlling for all other factors, age was positively associated with anaemia. Compared with women aged 15–24, older women aged 35–49 were 1.31 times more likely to be anaemic (aOR = 1.31; 95% CI 1.01–1.70). The results suggest that women who are currently married are 0.72 times as likely to have anaemia than those who are not in union (aOR = 0.72; 95% CI 0.28–0.78). Anaemia varies significantly with education. The results suggest that women's odds of having anaemia decrease with an increased level of completed education. Women with more than secondary education are associated with 0.47 times the odds of anaemia compared to those with primary or less education (aOR = 0.47; 95% CI 0.28–0.78). Household wealth is positively associated with anaemia. Compared to poor women, women of average or middle-class household wealth had 1.32 times higher odds of being anaemic (aOR = 1.32; 95% CI 1.03–1.69). Concerning occupational status, women in professional, technical, or clerical positions are 60% more likely to have anaemia compared to their unemployed counterparts (aOR = 1.60; 95% CI 1.04–2.47). Women who are currently using modern contraceptives are less likely to have anaemia than those who are not using them (aOR = 0.59; 95% CI 0.49–0.72). The results suggest that body mass index is negatively associated with anaemia. Women's likelihood of having anaemia decreased with an increase in body mass index. Overweight and obese women had 0.56 times the odds of having anaemia compared to underweight women (aOR = 0.56; 95% CI 0.37–0.86). Anaemia varies significantly with women's province of residence. The Province (or region) of residence is an important factor associated with anaemia in Zimbabwe. Relative to Manicaland province, Matabeleland South had the highest burden of anaemia. Midlands and Harare also had a fair share of high anaemia (Table 3).

Table 1 Percentage distribution of respondents by selected characteristics, ZDHS 2015 (N = 5412)

Variable	Frequency (%)	Variable	Frequency (%)
Age		Type of toilet facility	
15–24	1681 (29.6)	Improved	3448 (60.1)
25–34	2681 (50.8)	Unimproved	1964 (39.9)
35–49	1050 (19.7)	Type of cooking fuel	
Age at first birth		Low-polluting	1759 (27.4)
< 18	1394 (26.1)	High-polluting	3653 (72.6)
18+	4018 (73.9)	Body Mass Index	
Marital status		< 18.5	235 (4.2)
Not in union	846 (14.6)	18.5–24.99	3215 (60.7)
Married	4566 (85.4)	≥ 25	1962 (35.1)
Parity		Smoking	
1–2	2563 (44.6)	No	5393 (99.7)
3–5	2419 (46.8)	Yes	19 (0.3)
6+	430 (8.6)	Pregnant	
Religion		No	5124 (94.6)
None/other	694 (12.1)	Yes	288 (5.4)
Catholic	261 (4.7)	Breastfeeding	
Protestant	716 (13.2)	No	3238 (59.3)
Pentecostal	1273 (21.7)	Yes	2174 (40.7)
Apostolic	2468 (48.3)	Slept under a mosquito net	
Education		No	4533 (85.4)
Primary/less	1626 (32.3)	Yes	879 (14.6)
Secondary	3486 (62.6)	Ever terminated pregnancy	
Higher	300 (5.1)	No	4746 (87.2)
Household wealth		Yes	666 (12.8)
Poor	2058 (42.6)	Area	
Middle	859 (17.3)	Urban	2026 (31.4)
Rich	2495 (40.1)	Rural	3386 (68.6)
Occupation		Province/region	
Unemployed	2635 (47.4)	Manicaland	603 (13.9)
Agricultural/manual labour	670 (14.7)	Mashonaland Central	614 (10)
Sales/services	1551 (28.4)	Mashonaland East	484 (9.6)
Professional/technical/clerical	308 (5.3)	Mashonaland West	629 (13.2)
Other	248 (4.2)	Matabeleland North	466 (4.8)
Modern contraceptive use		Matabeleland South	432 (3.9)
No	1409 (26.3)	Midlands	609 (13.3)
Yes	4003 (73.7)	Masvingo	564 (12.1)
Access to media		Harare	619 (14.9)
No	1394 (27.1)	Bulawayo	392 (4.2)
Yes	4018 (72.9)	Anaemia	
Source of drinking water		No	4075 (75.9)
Improved	4279 (74.7)	Yes	1337 (24.1)
Unimproved	1133 (25.3)		

3.4 Performance evaluation of anaemia predictive models

The K-Nearest Neighbors, Decision Trees and random forest classifiers were used to predict anaemia among women of reproductive age. After applying anaemia predictive models (KNN, DT and RF), their performance was evaluated using accuracy, recall, F1-score, precision, area under the precision-recall curve(AUC-PR), Cohen's Kappa, Matthews

Table 2 Prevalence of anaemia among women by selected characteristics, ZDHS 2015

Variable	Anaemia (N = 5412)		P-value
	No (N = 4075)	Yes (N = 1337)	
Age			0.603
15–24	76.2	23.8	
25–34	76.4	23.6	
35–49	74.4	25.6	
Age at first birth			0.935
< 18	76.1	24.0	
18+	75.9	24.1	
Marital status			< 0.001***
Not in union	66.3	33.7	
Married	77.6	22.4	
Parity			0.165
1–2	75.4	24.7	
3–5	77.3	22.7	
6+	71.8	28.2	
Religion			0.237
None/other	73.2	26.8	
Catholic	69.4	30.6	
Protestant	76.0	24.1	
Pentecostal	76.1	23.9	
Apostolic	77.2	22.8	
Education			0.331
Primary/less	75.0	25.0	
Secondary	76.1	23.9	
Higher	80.3	19.7	
Household wealth			0.134
Poor	77.9	22.1	
Middle	74.0	26.0	
Rich	74.8	25.3	
Occupation			0.266
Unemployed	76.1	24.0	
Agricultural/manual labor	78.1	21.9	
Sales/services	76.3	23.7	
Professional/technical/clerical	71.3	28.7	
Other	70.8	29.2	
Modern contraceptive use			< 0.001***
No	67.7	32.3	
Yes	78.9	21.1	
Access to media			0.989
No	75.9	24.1	
Yes	76.0	24.1	
Source of drinking water			0.532
Improved	75.6	24.4	
Unimproved	76.9	23.1	
Type of toilet facility			0.859
Improved	75.8	24.2	
Unimproved	76.1	23.9	
Type of cooking fuel			0.103
Low-polluting	73.8	26.2	
High-polluting	76.7	23.3	

Table 2 (continued)

Variable	Anaemia (N = 5412)		P-value
	No (N = 4075)	Yes (N = 1337)	
Body Mass Index			0.042*
< 18.5	68.7	31.3	
18.5–24.99	75.3	24.7	
≥ 25	78.0	22.0	
Smoking			0.131
No	76.0	24.0	
Yes	54.0	46.0	
Pregnant			0.006**
No	76.5	23.5	
Yes	66.6	33.4	
Breastfeeding			0.058
No	74.6	25.4	
Yes	77.9	22.1	
Slept under a mosquito net			0.121
No	75.4	24.6	
Yes	79.0	21.0	
Ever terminated pregnancy			0.787
No	76.0	24.0	
Yes	75.4	24.6	
Area			0.121
Urban	73.9	26.1	
Rural	76.9	23.1	
Province/region			< 0.001***
Manicaland	83.8	16.2	
Mashonaland Central	75.4	24.7	
Mashonaland East	81.1	18.9	
Mashonaland West	77.1	23.0	
Matabeleland North	76.3	23.8	
Matabeleland South	59.0	41.0	
Midlands	71.8	28.2	
Masvingo	77.4	22.6	
Harare	71.7	28.3	
Bulawayo	75.4	24.6	

*P < 0.05, **P < 0.01, ***P < 0.001

correlation coefficient (MCC) and balanced accuracy, as shown in Table 4. Random forest classifier performed better than other models, with 74% accuracy, 78% recall, 75% F1-score, and a precision of 72%. However, the decision trees classifier also performed better than KNN, with 73% accuracy, 75% recall, 73% F1-score, and 71% precision.

Furthermore, the KNN model has the same low both Cohen's Kappa (36%) and MCC (36%) which suggests that the agreement between predicted and actual labels is weaker than the other models. Notably, Decision Trees' Cohen's Kappa and MCC are also higher (46%), indicating better reliability and agreement between predictions and actual values. This model appears more balanced in terms of both identifying anaemic and non-anaemic women correctly. Among these models, the Random Forest model recorded the highest F1-score (75%) and its Cohen's Kappa and MCC values (48%) show improved agreement between predictions and reality, indicating better overall performance and robustness. The balanced accuracy is the same as the overall accuracy, further confirming the model's effectiveness at handling imbalanced datasets and predicting anaemia among women of reproductive age (Fig. 4).

Furthermore, the performance of KNN, random forest, and decision tree was validated using StratifiedKFold validation to ensure that the class distribution in the dataset is preserved in each fold. Since the dataset had imbalanced classes, the StratifiedKFold validation technique became very useful in reducing the risk of biased evaluation. The performance

Table 3 Results from multivariate logistic regression analysis examining the association of having anaemia among the respondents

Variable	aOR (95% CI)
Age	
15–24	1
25–34	1.15 (0.93–1.43)
35–49	1.31 (1.01–1.70)*
Marital status	
Not in union	1
Married	0.72 (0.57–0.91)**
Education	
Primary/less	1
Secondary	0.86 (0.70–1.05)
Higher	0.47 (0.28–0.78)**
Household wealth	
Poor	1
Middle	1.32 (1.03–1.69)*
Rich	1.23 (0.91–1.67)
Occupation	
Unemployed	1
Agricultural/manual labour	0.94 (0.72–1.23)
Sales/services	0.97 (0.79–1.20)
Professional/technical/clerical	1.60 (1.04–2.47)*
Other	0.96 (0.64–1.45)
Modern contraceptive use	
No	1
Yes	0.59 (0.49–0.72)***
Body Mass Index	
< 18.5	1
18.5–24.99	0.72 (0.48–1.06)
≥ 25	0.56 (0.37–0.86)**
Province/region	
Manicaland	1
Mashonaland Central	1.82 (1.27–2.61)***
Mashonaland East	1.29 (0.86–1.96)
Mashonaland West	1.73 (1.13–2.65)*
Matabeleland North	1.67 (1.07–2.61)*
Matabeleland South	3.44 (2.32–5.09)***
Midlands	2.18 (1.50–3.17)***
Masvingo	1.55 (1.01–2.38)*
Harare	2.14 (1.43–3.20)***
Bulawayo	1.71 (1.08–2.71)*

*P < 0.05, **P < 0.01, ***P < 0.001

Table 4 Summary of performance across models

Predictive models	Accuracy (%)	Recall (%)	F1-score (%)	Precision (%)	AUC-PR (%)	Cohen's Kappa (%)	MCC (%)	Balanced accuracy (%)
KNN	68	74	70	66	68	36	36	68
DT	73	75	73	71	73	46	46	73
RF	74	78	75	72	74	48	48	74

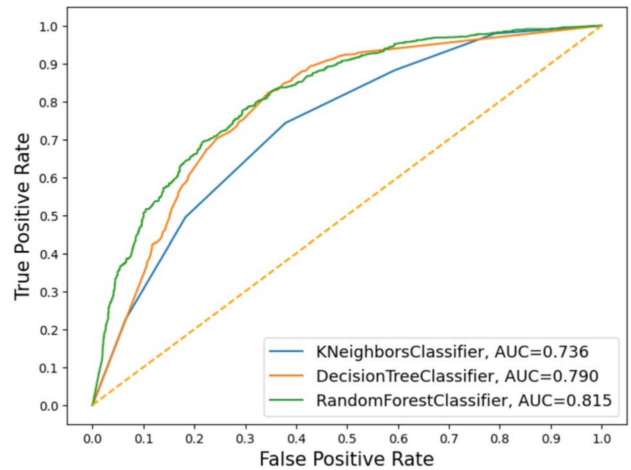

Fig. 4 Anaemia predictive models ROC AUC

of KNN and Decision Tree did not change using StratifiedKFold validation, as shown in Figs. 5 and 6, respectively. However, the performance of the random forest slightly improved with StratifiedKFold validation, as shown in Fig. 7. Again, the results show that the random forest is a better anaemia predictive model as compared to KNN and decision trees.

4 Discussion

This study examined the factors associated with anaemia in Zimbabwean women of reproductive age. The anaemia prevalence was 24.1% among women of reproductive age. This prevalence is classified as moderate public health significance and is similar to other studies within Africa and Zimbabwe [2, 7, 25]. This could be a result of nutritional deficiencies and various non-nutritional factors which we examined in this study. Our findings from the multivariate logistic analysis suggest that age, marital status, education, household wealth, occupation, modern contraceptive use, body mass index, and province/region were significantly associated with anaemia.

Previous studies have shown age is associated with anaemia [1, 12, 56]. This study revealed that women's age was positively associated with anaemia. Being older age (35–49) was associated with a higher likelihood of anaemia compared with those women of younger age (15–24). Similar findings have been found in many resource-poor countries [57], Eastern Africa [7, 19], and Myanmar [58]. Other studies have shown a not conclusive or unclear pattern in the relationship between age and being anaemic among women of reproductive age [3, 7]. Consistent with prior studies [18, 58] being

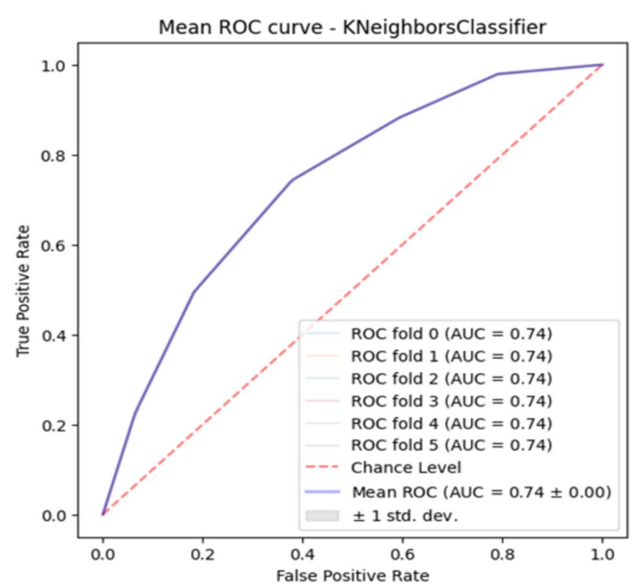

Fig. 5 KNN StratifiedKFold validation

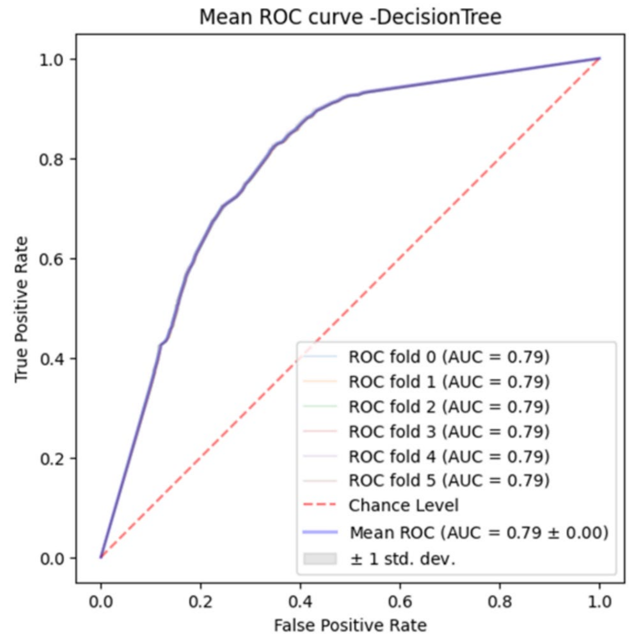

Fig. 6 Decision tree StratifiedKFold validation

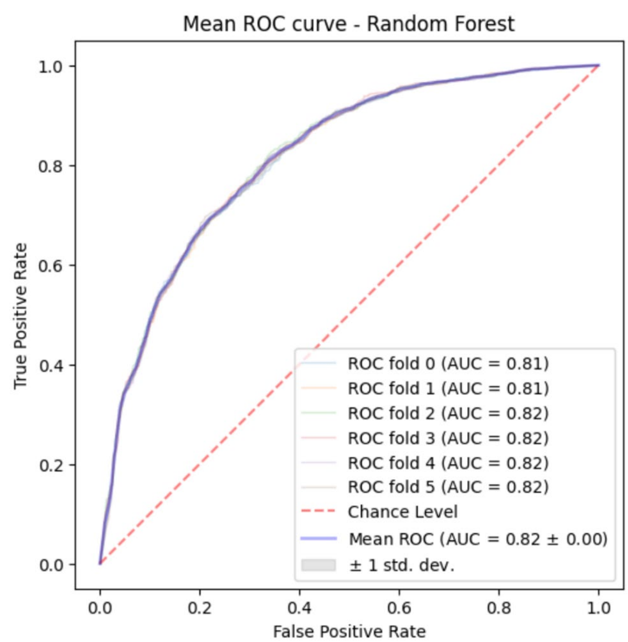

Fig. 7 Random forest StratifiedKFold Validation

currently married was associated with a lower likelihood of anaemia as compared with those women who were unmarried (never-married or previously married). Previously married women tend to be predisposed to economic hardships of malnutrition and poor health care increasing their likelihood of being anaemic [19]. Contrary, some studies reveal that being married significantly raised the likelihood of having anaemia among women of reproductive age [18, 20]. Education was negatively associated with anaemia. The likelihood of anaemia was lower among women who had higher education than those women who had primary or less education. This finding is quite consistent with previous research in sub-Saharan Africa [1, 7, 28, 32], South Asian countries [3], and among adolescent girls [21]. Being more educated may enable women to adopt health-seeking behaviours that can reduce their risk of anaemia by increasing their knowledge of health services, healthy eating, and hygiene practices [7, 19].

As with education, it is expected that women having a better economic status would tend to have improved diet, access to better health services, and living conditions tend to have a reduced risk of iron deficiency [19]. In contrast to earlier research that suggested there was a negative relationship being anaemia and wealth [10, 12, 26], our study found that household wealth was positively associated with anaemia. Women from households with middle or average wealth status are more likely to be anaemic compared to those from poor households. Similarly, women's occupation in professional, technical, or clerical occupations was more likely associated with having anaemia compared to women unemployed. In another study in Ethiopia, housewives and private workers were found to have a lower risk of anaemia [24]. Our study is in line with previous studies' expectations [26], women employed (except for women's professional job status) were less likely to have anaemia compared to those unemployed though the association in other occupations was not significant. In the study, women currently using modern contraceptives had a lower likelihood of anaemia than their counterparts not using contraceptives in line with previous studies [6, 26, 59, 60]. Studies have shown that modern contraceptives such as hormonal contraceptives lower the risk of anaemia by reducing bleeding or blood loss during menstruation although this evidence is contrary in some instances [19, 57]. From this study, an increase in the body mass index was found associated with lower odds of anaemia. Being overweight and obese was associated with a lower likelihood of anaemia as compared to those women who were underweight. This is in line with studies conducted in Nigeria [26], Ethiopia [12], Rwanda [19, 22], and Asian countries [3, 58]. Lack of important micronutrients such as iron, vitamins, and folate is the notable cause of anaemia in malnourished women [2, 4].

Our study findings showing a significant association between anaemia and region or province of residence are similar to past studies [19, 21, 25]. Overall, relative to Manicaland province, women residing in other specific provinces increased the likelihood of being anaemic. Women from Matabeleland South had the highest odds of having anaemia in line with a previous study in Zimbabwe [25]. Matabeleland province is an arid region most prone to droughts. The higher likelihood of developing anaemia may be linked to various issues including discrepancies in health care provision, health services access, and utilisation, disease risk, and healthy eating habits among the different population groups in Zimbabwe. Contrary to the literature we reviewed, the findings show that anaemia is not associated with factors such as age at first birth, parity, religion, access to the media, type of toilet facility, source of drinking water, type of cooking fuel, smoking, being pregnant, breastfeeding, sleeping under a mosquito net, pregnancy termination, and area or place of residence in women of reproductive age.

Furthermore, the Elastic Net indicates that contraceptive use, region, BMI and age are the most important features in predicting anaemia among women of reproductive age in Zimbabwe. This corroborates with a study by Teshome et al. [60] which highlighted that contraceptive use and BMI have a positive association with anaemia among women of reproductive age. Despite the dearth of literature, our study shows that there is an opportunity to apply machine learning models such as Random Forest, Decision Trees and KNN to predict anaemia among women of reproductive age with better accuracy. As evidenced in our findings, random forest predicted anaemia among women of reproductive age with better accuracy of 74%, recall of 78%, F1-Score of 75%, Precision of 72% and ROC AUC of 82% after applying StratifiedKFold validation. The random forest model outperformed the others, likely due to its ability to handle non-linear relationships, robust feature selection through ensemble methods and its inherent bootstrap aggregating (bagging) process. Random forest captures complex interactions between demographic and socioeconomic factors such as wealth, region, contraceptive use, and body mass index.

4.1 Implications of the study on policy and practice

The study findings present tremendous opportunities for healthcare professionals, public health experts and policymakers to develop effective interventions for early pre-anaemia screening that can be deployed in healthcare settings. Such machine learning models can be utilised together with existing interventions such as anaemia supplements to reduce the catastrophic impact of anaemia among women of reproductive age [43]. However, before drawing key policy recommendations, there is a need to deploy the random forest model in real health facilities by enrolling and collecting recent data (significant features) from women of reproductive age and further validating the performance of the model. This could be part of the future work. Once the desired performance threshold has been reached and the model validated, the early anaemia screening model can be presented to the relevant policymakers for consideration as a pre-anaemia screening tool for women of reproductive age. Once validated and accepted, it is important to incorporate machine learning predictive models in the country's health policies and guidelines and possibly integrate them into the existing electronic health records and health information systems.

4.2 Limitations of the study and future work

Despite our study based on nationally representative data of the Zimbabwe DHS, it has not been without its limitations. Responses to the questionnaire may be subjected to biases of respondents or interviewers. The cross-sectional nature of the study allows temporal but not causative relationships to be inferred from the findings. Excluding non-responses and missing cases in the exposure or outcome might introduce bias in our study results. Clinical infections and dietary intake known to impact anaemia prevalence were not investigated because such data was not collected by the Zimbabwe DHS and therefore did not form part of this study objective. The study used the 2015 ZDHS, therefore, there is a need to use more recent Zimbabwe DHS data when the data becomes available and accessible.

5 Conclusion

Overall, our findings indicate that anaemia among women of reproductive age is an important public health problem in Zimbabwe. Adjusted multivariate logistic regression results suggest that anaemia is significantly higher among older women (35–49 years), women with primary or less education, women not using modern contraceptives, underweight women, and women residing in all provinces except Mashonaland East. Provincial disparities in maternal anaemia were notable with the prevalence of anaemia lowest in Manicaland and highest in Matabeleland South. The Zimbabwe Ministry of Health and Child Care should actively consider scaling up anaemia screening and parasitic infection treatments for those women at the highest risk of being anaemic. Intervention strategies to ameliorate the impact of anaemia risk factors among women of reproductive age in Zimbabwe are iron and folic acid supplementation. Such low-cost interventions will need to be enhanced and prioritised in line with the SDGs of empowering women (goal 5), promoting health (goal 3), and improving dietary intake of iron-rich foods and other micronutrients (goal 2). Moreover, random forest machine learning model performed better than other anaemia predictive models in terms of accuracy, recall, F1-score, precision and AUC. Apply such model in developing data-driven early anaemia prediction tools that can assist in identifying risk factors associated with anaemia and predict anaemia. This can assist healthcare professionals, public health experts and policymakers to make informed decisions and design proactive measures and health interventions including optimizing the distribution and allocation of anaemia supplements to improve overall health outcomes.

Acknowledgements The authors are grateful to ZIMSTAT and Measure DHS program for granting permission to use the 2015 Zimbabwe Demographic and Health Survey. We extend our appreciation to the reviewers for their fruitful comments and suggestions.

Author contributions GBC: conceptualization, formal analysis, writing-original draft preparation. EM: conceptualization, formal analysis, writing review and editing, TD: writing, review and editing, GM: validation, writing-review and editing, BM: formal analysis, writing, review and editing SF: writing, review and editing, MM: software, writing-review and editing, RBM: validation, writing-review and editing. JB: software, writing and editing. EP: writing-review and editing.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability The datasets generated and/or analysed during the current study are available in the Measure DHS repository, http://www.dhsprogram.com.

Declarations

Ethics approval and consent to participate We used secondary data from the Zimbabwe National Statistics Agency (ZIMSTAT) which carried out the 2015 ZDHS survey. Therefore, this study did not require ethical approval since the data is available in the public domain upon request from the DHS measure program website (http://www.dhsprogram.com). The downloaded dataset is anonymised so that there are no individual identifiers reported in the data set. The adherence to ethical protocols and guidelines of consent, privacy, and confidentiality in the collection and archiving of data was ensured by the primary investigator (ZIMSTAT), producers, and funding agencies [29].

Competing interests The authors declare no competing interests.

References

1. Correa-Agudelo E, Kim H-Y, Musuka GN, Mukandavire Z, Miller FD, Tanser F, et al. The epidemiological landscape of anemia in women of reproductive age in sub-Saharan Africa. Sci Rep. 2021;11:11955. https://doi.org/10.1038/s41598-021-91198-z.
2. Stevens GA, Paciorek CJ, Flores-Urrutia MC, Borghi E, Namaste S, Wirth JP, et al. National, regional, and global estimates of anaemia by severity in women and children for 2000–19: a pooled analysis of population-representative data. Lancet Glob Health. 2022;10:e627–39. https://doi.org/10.1016/S2214-109X(22)00084-5.
3. Sunuwar DR, Singh DR, Chaudhary NK, Pradhan PMS, Rai P, Tiwari K. Prevalence and factors associated with anemia among women of reproductive age in seven South and Southeast Asian countries: evidence from nationally representative surveys. PLoS ONE. 2020;15: e0236449. https://doi.org/10.1371/journal.pone.0236449.
4. Armah-Ansah EK. Determinants of anemia among women of childbearing age: analysis of the 2018 Mali demographic and health survey. Arch Public Health. 2023;81:10. https://doi.org/10.1186/s13690-023-01023-4.
5. Young MF, Oaks BM, Tandon S, Martorell R, Dewey KG, Wendt AS. Maternal hemoglobin concentrations across pregnancy and maternal and child health: a systematic review and meta-analysis. Ann N Y Acad Sci. 2019;1450:47–68. https://doi.org/10.1111/nyas.14093.
6. Zegeye B, Anyiam FE, Ahinkorah BO, Ameyaw EK, Budu E, Seidu A-A, et al. Prevalence of anemia and its associated factors among married women in 19 sub-Saharan African countries. Arch Public Health. 2021;79:214. https://doi.org/10.1186/s13690-021-00733-x.
7. Teshale AB, Tesema GA, Worku MG, Yeshaw Y, Tessema ZT. Anemia and its associated factors among women of reproductive age in eastern Africa: a multilevel mixed-effects generalized linear model. PLoS ONE. 2020;15: e0238957. https://doi.org/10.1371/journal.pone.0238957.
8. Kinyoki D, Osgood-Zimmerman AE, Bhattacharjee NV, Lazzar-Atwood A, Lu D, Local Burden of Disease Anaemia Collaborators, et al. Anemia prevalence in women of reproductive age in low- and middle-income countries between 2000 and 2018. Nat Med. 2021;27:1761–82. https://doi.org/10.1038/s41591-021-01498-0.
9. Nti J, Afagbedzi S, da-Costa Vroom FB, Ibrahim NA, Guure C. Variations and determinants of anemia among reproductive age women in five sub-Saharan Africa countries. BioMed Res Int. 2021;2021:1–14. https://doi.org/10.1155/2021/9957160.
10. Abate TW, Getahun B, Birhan MM, Aknaw GM, Belay SA, Demeke D, et al. The urban–rural differential in the association between household wealth index and anemia among women in reproductive age in Ethiopia, 2016. BMC Womens Health. 2021;21:311. https://doi.org/10.1186/s12905-021-01461-8.
11. Chaparro CM, Suchdev PS. Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. Ann N Y Acad Sci. 2019. https://doi.org/10.1111/nyas.14092.
12. Ejigu BA, Wencheko E, Berhane K. Spatial pattern and determinants of anaemia in Ethiopia. PLoS ONE. 2018;13: e0197171. https://doi.org/10.1371/journal.pone.0197171.
13. Bhatia BK, Ratnani R. Prevalence of anemia in pregnant women attending CCM Medical College and Hospital Chhattisgarh. J Med Sci Clin Res. 2018. https://doi.org/10.18535/jmscr/v6i3.69.
14. Fondjo LA, Addai-Mensah O, Annani-Akollor ME, Quarshie JT, Boateng AA, Assafuah SE, et al. A multicenter study of the prevalence and risk factors of malaria and anemia among pregnant women at first antenatal care visit in Ghana. PLoS ONE. 2020;15: e0238077. https://doi.org/10.1371/journal.pone.0238077.
15. Brittenham GM, Moir-Meyer G, Abuga KM, Datta-Mitra A, Cerami C, Green R, et al. Biology of anemia: a public health perspective. J Nutr. 2023;153:S7-28. https://doi.org/10.1016/j.tjnut.2023.07.018.
16. Daru J, Zamora J, Fernández-Félix BM, Vogel J, Oladapo OT, Morisaki N, et al. Risk of maternal mortality in women with severe anaemia during pregnancy and post partum: a multilevel analysis. Lancet Glob Health. 2018;6:e548–54. https://doi.org/10.1016/S2214-109X(18)30078-0.
17. Mengistu GT, Mengistu BK, Gudeta TG, Terefe AB, Habtewold FM, Senbeta MD, et al. Magnitude and factors associated with iron supplementation among pregnant women in Southern and Eastern Regions of Ethiopia: further analysis of mini demographic and health survey 2019. BMC Nutr. 2022;8:66. https://doi.org/10.1186/s40795-022-00562-3.
18. Belay DG, Adane SM, Ferede OL, Lakew AM. Geographically weighted regression analysis of anemia and its associated factors among reproductive age women in Ethiopia using 2016 demographic and health survey. PLoS ONE. 2022;17: e0274995. https://doi.org/10.1371/journal.pone.0274995.
19. Hakizimana D, Nisingizwe MP, Logan J, Wong R. Identifying risk factors of anemia among women of reproductive age in Rwanda—a cross-sectional study using secondary data from the Rwanda demographic and health survey 2014/2015. BMC Public Health. 2019;19:1662. https://doi.org/10.1186/s12889-019-8019-z.
20. Worku MG, Tesema GA, Teshale AB. Prevalence and determinants of anemia among young (15–24 years) women in Ethiopia: a multilevel analysis of the 2016 Ethiopian demographic and health survey data. PLoS ONE. 2020;15: e0241342. https://doi.org/10.1371/journal.pone.0241342.
21. Atalell KA, Tamir TT, Alemu TG, Techane MA. Spatial distributions and determinants of anaemia among adolescent girls in Ethiopia: a secondary analysis of EDHS 2016 – a cross-sectional study. BMJ Open. 2022;12: e059405. https://doi.org/10.1136/bmjopen-2021-059405.
22. Habyarimana F, Zewotir T, Ramroop S. Spatial distribution and analysis of risk factors associated with anemia among women of reproductive age: case of 2014 Rwanda demographic and health survey data. Open Public Health J. 2018;11:425–37. https://doi.org/10.2174/1874944501811010425.
23. Lakew Y, Biadgilign S, Haile D. Anaemia prevalence and associated factors among lactating mothers in Ethiopia: evidence from the 2005 and 2011 demographic and health surveys. BMJ Open. 2015;5:e006001–e006001. https://doi.org/10.1136/bmjopen-2014-006001.

24. Mekonnen FA, Ambaw YA, Neri GT. Socio-economic determinants of anemia in pregnancy in North Shoa Zone, Ethiopia. PLoS ONE. 2018;13: e0202734. https://doi.org/10.1371/journal.pone.0202734.
25. Gona PN, Gona CM, Chikwasha V, Haruzivishe C, Mapoma CC, Rao SR. Intersection of HIV and Anemia in women of reproductive age: a 10-year analysis of three Zimbabwe demographic health surveys, 2005–2015. BMC Public Health. 2021;21:41. https://doi.org/10.1186/s12889-020-10033-8.
26. Ogbuabor DC, Ogbuabor AO, Ghasi N. Determinants of anaemia prevalence in women of reproductive age in Nigeria: a cross-sectional study using secondary data from Nigeria Demographic and Health Survey 2018. Womens Health. 2022;18:1–24. https://doi.org/10.1177/17455057221142961.
27. Kanno GG, Geremew T, Diro T, Musarapasi SV, Wyk RV, Seboka BT, et al. The link between indoor air pollution from cooking fuels and anemia status among non-pregnant women of reproductive age in Ethiopia. SAGE Open Med. 2022;10:1–9. https://doi.org/10.1177/20503121221107466.
28. 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: e027276. https://doi.org/10.1136/bmjopen-2018-027276.
29. Zimbabwe National Statistics Agency, ICF International. Zimbabwe Demographic and Health Survey 2015: Final Report. Rockville, Maryland, USA: Zimbabwe National Statistics Agency (ZIMSTAT) and ICF International; 2016.
30. Liyew AM, Tesema GA, Alamneh TS, Worku MG, Teshale AB, Alem AZ, et al. Prevalence and determinants of anemia among pregnant women in East Africa; a multi-level analysis of recent Demographic and Health Surveys. PLoS ONE. 2021;16: e0250560. https://doi.org/10.1371/journal.pone.0250560.
31. Weze K, Abioye AI, Obiajunwa C, Omotayo M. Spatio-temporal trends in anaemia among pregnant women, adolescents and preschool children in sub-Saharan Africa. Public Health Nutr. 2021;24:3648–61. https://doi.org/10.1017/S1368980020004620.
32. Tusa BS, Weldesenbet AB, Bahiru N, Enyew DB. Magnitudes of anemia and its determinant factors among lactating mothers in east African countries: using the generalized mixed-effect model. Front Nutr. 2021;8:1–8. https://doi.org/10.3389/fnut.2021.667466.
33. Chigariro TC, Mhloyi MM. Emigration as a social and economic determinant of health in low-income urban Zimbabwe. J Health Popul Nutr. 2022;41:49. https://doi.org/10.1186/s41043-022-00330-w.
34. Mugandani R, Wuta M, Makarau A, Chipindu B. Re-classification of agro-ecological regions of Zimbabwe in conformity with climate variability and change. Afr Crop Sci J. 2012;20:361–9.
35. Rutstein S, Johnson K. The DHS wealth index. Calverton: ORC Macro; 2004.
36. Islam MDM, Rahman MDJ, Islam MM, Roy DC, Ahmed NAMF, Hussain S, et al. Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh. Int J Cogn Comput Eng. 2022;3:46–57. https://doi.org/10.1016/j.ijcce.2022.02.002.
37. Dzinamarira T, Mbunge E, Chingombe I, Cuadros DF, Moyo E, Chitungo I, et al. Using machine learning models to plan HIV services: emerging opportunities in design, implementation and evaluation. SAMJ South Afr Med J. 2024;114:86–90. https://doi.org/10.7196/SAMJ.2024.v114i6b.1439.
38. Mbunge E, Ndumiso N, Kavu TD, Dandajena K, Batani J, Fashoto SG. Towards QR code health systems amid COVID-19: lessons learnt from other QR code digital technologies. In: Adibi S, Rajabifard A, Shariful Islam SM, Ahmadvand A, editors. Science behind the COVID pandemic and healthcare technology solution. Cham: Springer International Publishing; 2022. p. 129–43. https://doi.org/10.1007/978-3-031-10031-4_7.
39. Mbunge E, Millham RC, Sibiya MN, Takavarasha S. Diverging mobile technology's cognitive techniques into tackling malaria in sub-Saharan Africa: a review. In: Silhavy R, Silhavy P, Prokopova Z, editors. Software engineering application in informatics. Cham: Springer International Publishing; 2021. p. 679–99. https://doi.org/10.1007/978-3-030-90318-3_54.
40. Chingombe I, Musuka G, Mbunge E, Chemhaka G, Cuadros DF, Murewanhema G, et al. Predicting HIV status using machine learning techniques and bio-behavioural data from the Zimbabwe population-based HIV impact assessment (ZIMPHIA15-16). In: Silhavy R, editor., et al., Artificial intelligence trends in system. Cham: Springer International Publishing; 2022. p. 247–58. https://doi.org/10.1007/978-3-031-09076-9_24.
41. Mbunge E, Sibiya MN. Mobile health interventions for improving maternal and child health outcomes in South Africa: a systematic review. Glob Health J. 2024;8:103–12. https://doi.org/10.1016/j.glohj.2024.08.002.
42. Meitei AJ, Saini A, Mohapatra BB, Singh KhJ. Predicting child anaemia in the North-Eastern states of India: a machine learning approach. Int J Syst Assur Eng Manag. 2022;13:2949–62. https://doi.org/10.1007/s13198-022-01765-4.
43. Vohra R, Hussain A, Dudyala AK, Pahareeya J, Khan W. Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting. PLoS ONE. 2022;17: e0269685. https://doi.org/10.1371/journal.pone.0269685.
44. Jana A, Chattopadhyay A, Saha UR. Identifying risk factors in explaining women's anaemia in limited resource areas: evidence from West Bengal of India and Bangladesh. BMC Public Health. 2022;22:1433. https://doi.org/10.1186/s12889-022-13806-5.
45. Appiahene P, Asare JW, Donkoh ET, Dimauro G, Maglietta R. Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. BioData Min. 2023;16:2. https://doi.org/10.1186/s13040-023-00319-z.
46. Dejene BE, Abuhay TM, Bogale DS. Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm. BMC Med Inform Decis Mak. 2022;22:247. https://doi.org/10.1186/s12911-022-01992-6.
47. Mbunge E, Chemhaka G, Batani J, Gurajena C, Dzinamarira T, Musuka G, et al. Predicting diarrhoea among children under five years using machine learning techniques. In: Silhavy R, editor., et al., Artificial intelligence trends in system. Cham: Springer International Publishing; 2022. p. 94–109. https://doi.org/10.1007/978-3-031-09076-9_9.
48. Mbunge E, Fashoto SG, Muchemwa B, Millham RC, Chemhaka G, Sibiya MN, et al. Application of machine learning techniques for predicting child mortality and identifying associated risk factors. In: 2023 conference on information communications and technology society. ICTAS, 2023, p. 1–5. https://doi.org/10.1109/ICTAS56421.2023.10082734.
49. Barus OP, Happy J, Jusin, Pangaribuan JJ, H SZ, Nadjar F. Liver Disease Prediction Using Support Vector Machine and Logistic Regression Model with Combination of PCA and SMOTE. In: 2022 1st international conference on technology innovation and its applications. ICTIIA, 2022, p. 1–6. https://doi.org/10.1109/ICTIIA54654.2022.9935879.

50. Mbunge E, Sibiya MN, Takavarasha S, Millham RC, Chemhaka G, Muchemwa B, et al. Implementation of ensemble machine learning classifiers to predict diarrhoea with SMOTEENN, SMOTE, and SMOTETomek class imbalance approaches. In: 2023 conference on information communication technology and society. ICTAS, 2023, p. 1–6. https://doi.org/10.1109/ICTAS56421.2023.10082744.
51. De Mol C, De Vito E, Rosasco L. Elastic-net regularization in learning theory. J Complex. 2009;25:201–30. https://doi.org/10.1016/j.jco.2009.01.002.
52. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005;67:301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.
53. Bansal M, Goyal A, Choudhary A. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decis Anal J. 2022;3: 100071. https://doi.org/10.1016/j.dajour.2022.100071.
54. Mishra S, Mallick PK, Tripathy HK, Bhoi AK, González-Briones A. Performance evaluation of a proposed machine learning model for chronic disease datasets using an integrated attribute evaluator and an improved decision tree classifier. Appl Sci. 2020;10:8137. https://doi.org/10.3390/app10228137.
55. Liu Y, Wang Y, Zhang J. New machine learning algorithm: random forest. In: Liu B, Ma M, Chang J, editors. Information computing and application. vol. 7473, Berlin, Heidelberg: Springer Berlin Heidelberg; 2012, p. 246–52. https://doi.org/10.1007/978-3-642-34062-8_32.
56. Nankinga O, Aguta D. Determinants of anemia among women in Uganda: further analysis of the Uganda demographic and health surveys. BMC Public Health. 2019;19:1757. https://doi.org/10.1186/s12889-019-8114-1.
57. Bellizzi S, Ali MM. Effect of oral contraception on anemia in 12 low- and middle-income countries. Contraception. 2018;97:236–42. https://doi.org/10.1016/j.contraception.2017.11.001.
58. Win H, Ko M. Geographical disparities and determinants of anaemia among women of reproductive age in Myanmar: analysis of the 2015–2016 Myanmar Demographic and Health Survey. WHO South-East Asia J Public Health. 2018;7:107–13. https://doi.org/10.4103/2224-3151.239422.
59. Gautam S, Min H, Kim H, Jeong H-S. Determining factors for the prevalence of anemia in women of reproductive age in Nepal: evidence from recent national survey data. PLoS ONE. 2019;14: e0218288. https://doi.org/10.1371/journal.pone.0218288.
60. Teshome AA, Berra WG, Hiruy AF. Modern contraceptive methods predict hemoglobin levels among women of childbearing age from DHS 2016. Open Access J Contracept. 2022;13:1–8. https://doi.org/10.2147/OAJC.S329045.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.