Socioeconomic, Nutritional, and Demographic Determinants of Anaemia Among Nigerian Women: A Machine Learning Analysis of DHS 2024 Data

Bunmi Segun Rotimi

Department of Mathematical Sciences

Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria

DOI: https://dx.doi.org/10.51584/IJRIAS.2025.10120081

Received: 17 December 2025; Accepted: 22 December 2025; Published: 17 January 2026

ABSTRACT

Anaemia remains a serious public health problem among women of reproductive age in Nigeria, with significant implications for maternal and population health. This study examined the prevalence, determinants, and predictability of anaemia using descriptive statistics, inferential analyses, and supervised machine-learning models on the data from the 2023–2024 Nigeria Demographic and Health Survey (NDHS). Anaemia status was defined using World Health Organisation haemoglobin thresholds. The results indicate an alarming 78% prevalence of anaemia among Nigerian women across geographic, socioeconomic, and educational strata. Nutritional status, particularly body mass index, together with reproductive factors and contextual characteristics, emerged as the most influential predictors of anaemia. The study highlights the value of combining epidemiological analysis with interpretable machine learning to inform targeted strategies for anaemia prevention and control in Nigeria.

Keywords: Anaemia; Women of reproductive age; Nigeria Demographic and Health Survey; Machine learning; Explainable artificial intelligence; Nutritional status; Public health surveillance

INTRODUCTION

Anaemia remains one of the most pressing global health concerns, particularly among women of reproductive age in low- and middle-income countries. The World Health Organisation (2023) estimates that over 30 per cent of women aged 15–49 years worldwide are anaemic, with sub-Saharan Africa contributing the largest share of cases. In Nigeria, despite substantial investments in maternal and nutritional health programmes, the prevalence of anaemia remains well above the WHO threshold for a severe public health problem (Ogbaju et al., 2025). The condition contributes to a considerable burden of maternal morbidity and mortality, adverse pregnancy outcomes, and diminished economic productivity (Gunarathna et al., 2024). According to Abioye et al., (2024), the causes of anaemia among Nigerian women are multifactorial, combining nutritional deficiencies, socioeconomic deprivation and infectious diseases.

Recent studies have provided updated insights into the patterns and drivers of anaemia in Nigeria and similar settings. Obeagu and Agreen (2023) analysed national survey data, and they reported that approximately 61 per cent of pregnant women were anaemic, with the highest risk among those with low educational attainment, short birth intervals, and poverty. Using spatial and multilevel modelling, Babah et al. (2025) identified significant regional clustering of anaemia across Nigeria, strongly associated with wealth index, educational level, and access to media information, implying that socio-economic context shapes both awareness and nutritional practices. Sanni (2025) further observed a growing burden of anaemia among urban and middle-income women, suggesting that changing dietary habits and the double burden of malnutrition are emerging challenges beyond classical poverty-related explanations. Comparable findings across sub-Saharan Africa by Tilahun et al. (2024) demonstrated that education, reproductive history, and household wealth consistently predict anaemia severity, while geographic inequities persist even after accounting for individual factors.

While these epidemiological analyses have advanced understanding of risk factors, most rely on conventional statistical models that assume linear relationships and limited interaction effects. Such methods may not adequately capture the complex, nonlinear interplay among demographic, nutritional, and socioeconomic variables. To address this limitation, recent research has begun applying machine-learning techniques to population health data. Zemariam et al. (2024) employed supervised learning algorithms to predict anaemia among adolescent girls in Ethiopia, utilising Demographic and Health Survey datasets, and achieved higher predictive accuracy than logistic regression models. Similarly, Kitaw et al., (2024) employed gradient-boosting algorithms to grade the severity of anaemia among pregnant women and utilised SHapley Additive exPlanations (SHAP) to interpret model outputs, identifying the relative influence of diet, education, and parity on risk classification. These advances demonstrate that machine learning can reveal complex, multidimensional determinants of anaemia and support more targeted intervention design.

Despite the growing global interest in data-driven methods, there remains a paucity of machine-learning research focused specifically on Nigerian women using the latest Demographic and Health Survey (DHS-VIII, 2023–2024) data. The availability of biomarker, anthropometric, and extensive socioeconomic indicators in this survey presents a unique opportunity to apply advanced analytical approaches to a nationally representative dataset. Addressing this gap is crucial because effective public-health programming requires not only quantifying anaemia prevalence but also identifying which combinations of nutritional, demographic, and contextual factors most strongly predict vulnerability.

This study, therefore, seeks to apply and evaluate machine-learning models, specifically logistic regression, random forest, gradient boosting, extreme gradient boosting (XGBoost), categorical boosting (CatBoost), and support vector machines, to predict anaemia among Nigerian women of reproductive age using DHS 2023–2024 data. In doing so, the research aims to achieve three main objectives: first, to examine the prevalence and distribution of anaemia across sociodemographic subgroups of women of reproductive age in Nigeria; second, to identify the most significant socioeconomic, nutritional, and reproductive determinants of anaemia using interpretable machine-learning method such as LIME; and third, to assess the predictive performance and practical utility of these models for informing public-health decision-making. This study contributes to evidence-based policy formulation aligned with Sustainable Development Goal 3 on ensuring healthy lives and promoting well-being for all by integrating modern computational techniques with nationally representative data. The study also demonstrates the role of artificial intelligence in advancing equitable maternal and nutritional health outcomes in Nigeria.

Research Methodology

This study employed a quantitative, analytical, and data-driven design to examine the socioeconomic, nutritional, and contextual determinants of anaemia among Nigerian women of reproductive age (WRA). Data were obtained from the 2023–2024 Nigeria Demographic and Health Survey (NDHS), implemented by the National Population Commission in collaboration with ICF under the DHS Program (NDHS, 2024). The NDHS uses a two-stage stratified cluster sampling design to ensure national representativeness across Nigeria's six geopolitical zones, first selecting enumeration areas as primary sampling units and subsequently households within each cluster (ICF, 2024). Figure 1 below is the description of the methodological pipeline employed in the study.

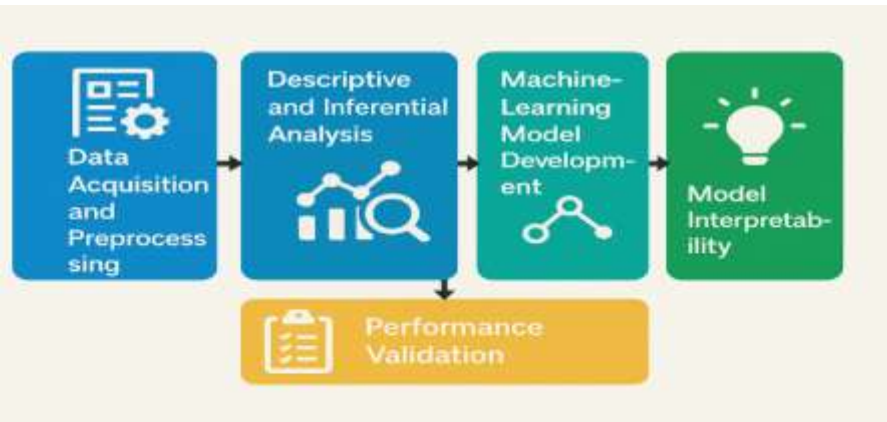

Figure 1. Methodology Pipeline (Source: author's design)

Study Population and Data Integration

The analytical sample comprised women aged 15–49 years who were de facto household members and had valid haemoglobin measurements. Two DHS recode files were integrated: the Household Members Recode (PR), which contains biomarker data including haemoglobin concentration, and the Individual Women's Recode (IR), which provides detailed socioeconomic, demographic, and reproductive health information. Datasets were merged using unique DHS identifiers, cluster number, household number, and line number, ensuring one-to-one correspondence between respondents. The resulting variables are shown in Table 1 below

Table 1. Description of Variables Used in the Analysis

Variable Name	Description	Measurement / Coding	Type
cluster	Primary sampling unit (enumeration area) identifier	DHS cluster code	Categorical (ID)
household	Household identification number within cluster	DHS household code	Categorical (ID)
line	Line number of respondent within household	DHS line number	Categorical (ID)
Anaemia	Anaemia status of the respondent	1 = Anaemic, 0 = Non-anaemic (based on WHO Hb thresholds)	Binary (Outcome)
Weight (kg)	Body weight of respondent	Measured in kilograms	Continuous
Height	Height of respondent	Measured in centimetres	Continuous
BMI	Body Mass Index	Weight (kg) / height^2 (m^2)	Continuous
Pregnancy	Pregnancy status at time of survey	1 = Pregnant, 0 = Not pregnant	Binary
Age	Age of respondent	Completed years (15–49)	Continuous
Region	Geopolitical zone of residence	1–6 (North West, North East, North Central, South East, South South, South West)	Categorical
Residence_Type	Type of place of residence	Urban / Rural (DHS coding)	Categorical
Religion	Religious affiliation	DHS religion categories	Categorical
Ethnicity	Ethnic group of the respondents	DHS ethnicity codes	Categorical
Education	Highest educational attainment	None, Primary, Secondary, Higher	Ordinal
Wealth_index	Household wealth quintile	Poorest–Richest (1–5)	Ordinal
Household_size	Number of household members	Count	Continuous
Birth_5yrs	Number of births in the last five years	Count	Continuous

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XII December 2025

Contraceptive_use	Current contraceptive method use	DHS contraceptive categories	Categorical
News_Paper	Frequency of reading newspapers	Not at all / Less than weekly / At least weekly	Ordinal
Radio	Frequency of listening to radio	Not at all / Less than weekly / At least weekly	Ordinal
Television	Frequency of watching television	Not at all / Less than weekly / At least weekly	Ordinal
Health_Insurance	Health insurance coverage	1 = Covered, 0 = Not covered	Binary
Total_children	Total number of children ever born	Count	Continuous

Outcome

Anaemia status was the primary outcome. Haemoglobin concentrations measured using HemoCue devices were converted from DHS raw values and classified according to World Health Organisation guidelines. Anaemia was defined as haemoglobin <12 g/dL for non-pregnant women and <11 g/dL for pregnant women (WHO, 2023). A binary outcome variable was constructed (1 = anaemic; 0 = non-anaemic). Anaemia severity categories were additionally derived for descriptive purposes.

Explanatory

Independent variables were selected based on established literature (Worku et al., 2022, Tirore et al., 2024, Tilahun et al., 2024) and grouped into conceptual domains: demographic characteristics (age, region, residence type, religion, ethnicity); socioeconomic status (education, household wealth index, household size); nutritional indicators (body mass index); reproductive and maternal factors (pregnancy status, births in the past five years, contraceptive use, total number of children ever born); and access to information and healthcare (media exposure and health insurance coverage). These variables capture both proximate and distal determinants of anaemia risk (Tirore et al., 2024).

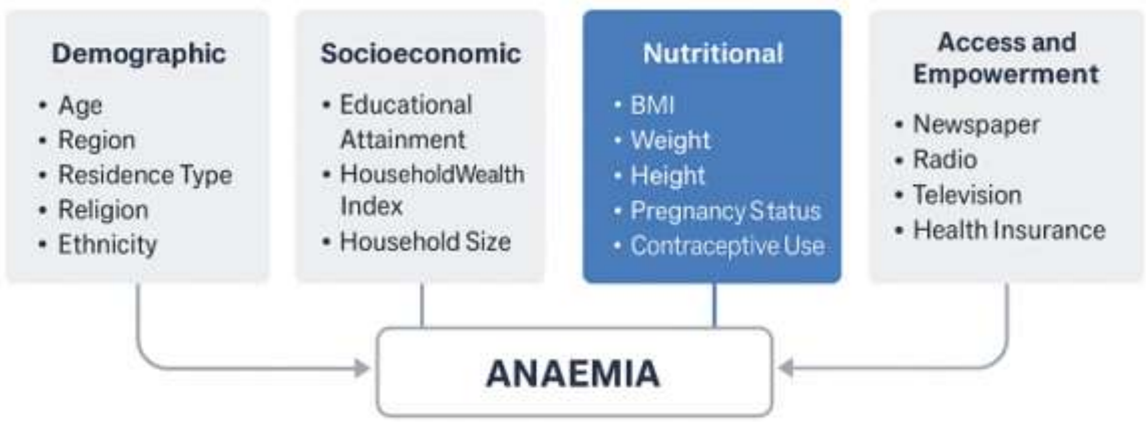

Figure 2. Conceptual framework illustrating the relationships between risk factors and anaemia among Nigerian women of reproductive age (Source: author's design).

Data Preprocessing

Data preprocessing involved systematic cleaning and transformation of the data in preparation for analysis and modelling (Cetin and Yildiz, 2022). Missing values were assessed and handled using median imputation for continuous variables and modal imputation for categorical variables, with missingness below five per cent for all key predictors. Continuous variables were standardised to improve model convergence. To minimise spatial leakage inherent in clustered survey data (Foster et al., 2024), a group-aware train–test split was applied, ensuring that all respondents within the same sampling cluster were assigned to the same partition. 80% of the data was used for training and 20% for testing. Given the high prevalence of anaemia, class imbalance was addressed through class-weighting during model training rather than synthetic oversampling (Ren et al., 2023).

Statistical Analysis

Descriptive statistics were computed to estimate anaemia prevalence and examine subgroup distributions across demographic and socioeconomic categories. Associations between categorical variables and anaemia status were assessed using chi-square tests of independence. Differences in continuous variables between anaemic and non-anaemic women were evaluated using one-way analysis of variance (ANOVA). These analyses provided epidemiological context and informed model specification.

Machine-Learning Modelling

Multiple supervised machine-learning algorithms were implemented to predict anaemia status, including logistic regression, random forest, gradient boosting, extreme gradient boosting (XGBoost), categorical boosting (CatBoost), and support vector machines. These models represent both linear and non-linear classifiers capable of capturing complex interactions among predictors. Models were trained on the training dataset and evaluated on the held-out test set. Performance metrics such as precision, recall, F1-score, accuracy, and the area under the receiver operating characteristic curve (ROC–AUC) (Obi, 2023) were used to evaluate the models.

Model Interpretability

To address the black-box nature of the machine learning models, explainable artificial intelligence (XAI) techniques were employed. Permutation feature importance was used as the primary method for ranking predictors based on their impact on model discrimination, providing unbiased, model-agnostic importance estimates (Molnar et al., 2024; Khan et al., 2025). Local explanations were further explored using Local Interpretable Model-Agnostic Explanations (LIME) (Zafar and Khan, 2021). LIME was applied to explain individual-level predictions, enhancing transparency and policy relevance.

Analytical Tools

All analyses were conducted using Python (version 3.12). Data management and statistical analyses utilised pandas, NumPy, and SciPy. Machine-learning models were implemented using scikit-learn, XGBoost, and CatBoost libraries, while model interpretability was done using the LIME library. Visualisations were generated using matplotlib and seaborn.

RESULTS AND DISCUSSION

This section presents and interprets the findings of the study. The results progress from descriptive analyses of anaemia burden and subgroup variations to inferential statistical tests. Machine-learning models are then applied to evaluate the joint predictive capacity of these variables, with performance assessed using standard classification metrics. To enhance transparency, explainable artificial intelligence techniques are employed to identify global and individual-level drivers of anaemia risk.

Descriptive Analysis

Anaemia remains a severe public-health burden among Nigerian women of reproductive age, with a prevalence of 78.4% (Figure 3), far exceeding the World Health Organisation threshold for a condition of major public-health significance (WHO, 2023). This high prevalence indicates that anaemia is widespread across the female population rather than concentrated in specific vulnerable subgroups. The finding is consistent with the result of Tirore et al., 2024; and Abubakar et al., 2024, which attributes persistently high anaemia rates in Sub-Saharan

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XII December 2025

Africa to a combination of nutritional inadequacies, malaria and helminth infections, and inflammation-related anaemia of chronic disease.

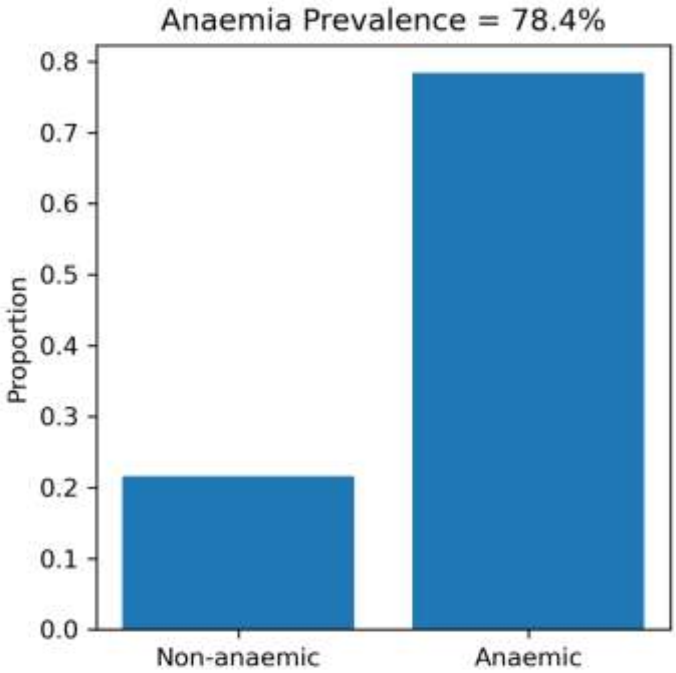

Figure 3. Anaemia Prevalence among Nigerian Women

Table 2 shows that anaemia in Nigeria is universal rather than selective, driven by a complex intersection of dietary patterns, socioeconomic transitions, infectious disease burden, and lifestyle factors. A significant regional variation was observed, with the North Central (83.1%) and South West (82.8%) recording the highest burden. Although northern zones often experience poorer maternal health outcomes, the unexpectedly high prevalence in southern zones may reflect the increasing double burden of malnutrition, particularly diets high in carbohydrates and low in micronutrient-rich foods, as well as urban lifestyle changes (Abioye et al., 2025; Aigbedion et al., 2025). Conversely, the South East had the lowest prevalence (73.0%), reinforcing findings from Ogunsakin et al. (2021) that cultural dietary patterns and differences in healthcare utilisation may influence regional disparities in anaemia.

However, patterns across socioeconomic variables reveal striking and counterintuitive trends. Anaemia prevalence increases with education and wealth, reaching 83.4% among the richest women and 83.1% among those with higher education. This contradicts traditional assumptions but reflects emerging evidence that urban, wealthy, and highly educated women in Africa may have poorer micronutrient intake due to reliance on processed foods and sedentary lifestyles (Sanni, 2025). Additionally, high prevalence among advantaged groups may also reflect inflammation-related anaemia associated with obesity, which is rising in Nigeria's urban centres (Babah et al., 2025).

Religious affiliation also showed strong variation, with Traditionalist (90.7%) and "Other religions" (100%) reporting the highest prevalence. Although these groups may represent small samples, the pattern could indicate disparities in healthcare access, cultural dietary restrictions, or reduced engagement with preventive services such as iron supplementation.

Table 2. Descriptive analysis of anaemia prevalence and subgroup distributions

Variable	**Category**	**Non-Anaemic (%)**	**Anaemic (%)**
Region	North West	23.8	76.2

	North East	24.4	75.6
	North Central	16.9	83.1
	South East	27.0	73.0
	South South	20.3	79.7
	South West	17.2	82.8
Education	None	26.2	73.8
	Primary	22.4	77.6
	Secondary	19.9	80.1
	Higher	16.9	83.1
Wealth index	Poorest	26.8	73.2
	Poorer	24.7	75.3
	Middle	22.5	77.5
	Richer	19.9	80.1
	Richest	16.6	83.4
Religion	Catholic	20.7	79.3
	Other Christians	19.7	80.3
	Islam	23.5	76.5
	Traditionalist	9.3	90.7
	Other religions	0.0	100.0

Associations Between Sociodemographic Factors and Anaemia Status

Anaemia status among Nigerian women of reproductive age is significantly associated with sociodemographic and healthcare-related variables (see chi-square results in Table 3). Table 3 reveals a strong association between region and anaemia (χ^2=112.64, $p<10^{-22}$), which indicates substantial geographic variation in disease burden. This aligns with findings from Abioye et al. (2025), who reported pronounced regional disparities in anaemia across Nigeria linked to malaria transmission patterns, access to fortified foods, and differences in dietary diversity. Also, education (χ^2=92.22, $p<10^{-20}$) and wealth index (χ^2=106.68, $p<10^{-22}$) are strongly associated with anaemia. According to Sanni (2025), women with lower education and those in poorer households traditionally face higher nutritional deficiencies, limited access to iron-rich diets, and greater exposure to parasitic infections. However, the descriptive analysis in Table 1 shows that anaemia is high among wealthier and more educated women, reflecting the emerging “nutrition transition” in Nigeria, where urban and affluent populations increasingly consume calorie-dense but micronutrient-poor foods (Aigbedion et al., 2025). These findings suggest that socioeconomic status alone is not sufficient to protect against anaemia, and that nutritional quality, rather than economic wealth, is a critical determinant.

Moreover, cultural factors are also important. The significant associations for religion ($p<10^{-7}$) and ethnicity ($p<10^{-15}$) indicate that health behaviours, dietary preferences, and traditional beliefs shape women’s nutritional

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XII December 2025

status (Bello et al., 2025). Furthermore, media exposure, including newspapers, radio, and television, was a significant predictor, reflecting the importance of access to health information. Women with limited media exposure may have lower nutrition literacy, reduced awareness of iron supplementation, and less knowledge about malaria prevention (Shatilwe et al., 2021). Finally, health insurance coverage ($p<0.001$) was significantly associated with anaemia, highlighting the role of financial access to healthcare. Insured women may be more likely to use antenatal care services, receive iron supplements, or obtain timely treatment for infections.

Table 3. Showing Association tests between sociodemographic variables and anaemia

Variable	χ^2	df	p-value	Significance
Region	112.64	5	1.13×10^{-22}	Significant
Religion	34.81	4	5.08×10^{-7}	Significant
Ethnicity	451.40	242	8.01×10^{-15}	Significant
Education	92.22	3	7.29×10^{-20}	Significant
Wealth index	106.68	4	3.72×10^{-22}	Significant
Contraceptive use	86.72	15	4.04×10^{-12}	Significant
Newspaper	22.54	2	1.28×10^{-5}	Significant
Radio	22.54	2	1.28×10^{-5}	Significant
Television	28.41	2	6.79×10^{-7}	Significant
Health insurance	11.30	1	7.76×10^{-4}	Significant

Differences in Continuous Sociodemographic and Nutritional Factors by Anaemia Status

The results in Table 4 assess whether selected continuous variables differ significantly between anaemic and non-anaemic women. The findings reveal that three variables, BMI, household size, and total number of children, show statistically significant differences across anaemia status, while age and recent births do not. The lack of a significant age difference ($F=0.00$, $p=0.970$) suggests that anaemia affects Nigerian women across the reproductive lifespan in a relatively uniform manner. This aligns with evidence from Wallace (2020), who noted that in high-burden settings, structural factors such as diet, infection, and socioeconomic conditions overshadow age-related physiological differences. The lack of a significant association for births within the last five years ($p=0.0815$) also indicates that recent childbirth alone does not explain current anaemia levels. This may reflect improved postpartum care or the dominance of other chronic determinants, such as diet and infection.

However, BMI shows a strong and highly significant association ($F=55.28$, $p<1\times10^{-13}$), indicating that nutritional status plays a central role in anaemia risk. Both underweight and overweight women may be vulnerable: undernutrition leads to iron deficiency, while overweight women are more prone to inflammation-mediated anaemia of chronic disease (Aboagye et al., 2023; Abubakar, 2024). Household size also demonstrates a significant difference ($F=12.34$, $p<0.001$), suggesting that women living in larger households may experience greater nutritional stress due to food insecurity or resource dilution. Larger families often allocate limited resources across more members, reducing dietary diversity and the likelihood of consuming iron-rich foods (Worku, 2022). This reinforces the importance of household economic pressures as underlying contributors to anaemia.

Similarly, the total number of children (an indicator of cumulative reproductive burden) is significantly associated with anaemia ($F=12.34$, $p<0.001$). High parity is a known risk factor for iron depletion, particularly in settings where postpartum supplementation uptake is low (Tirore, 2024). The significance of lifetime parity, rather than recent births, suggests that long-term reproductive history, rather than short-term fertility, has a greater influence on anaemia status.

Table 4. Testing Significant difference in sociodemographic continuous variables and anaemia status

Variable	F-Value	p-value	Significance
Age	0.00	0.970	Not significant
BMI	55.28	1.11×10^{-13}	Significant
Household size	12.34	4.44×10^{-4}	Significant
Births in 5 yrs	3.04	0.0815	Not Significant
Total children	12.34	4.44×10^{-4}	Significant

Machine-Learning Model Performance for Anaemia Prediction

Table 5 indicates that all machine-learning models achieved relatively high precision and F1-scores for predicting anaemia. Gradient Boosting and Support Vector Machine models achieved perfect recall (1.00), indicating that they successfully detected all anaemic cases (Obi, 2023). Random Forest, XGBoost, and CatBoost showed more balanced performance, with moderate recall and competitive F1-scores, reflecting a trade-off between sensitivity and precision. The comparatively lower recall in Logistic Regression highlights the limitations of linear models in capturing complex, non-linear relationships inherent in population health data (Zemariam et al., 2024). The dominance of tree-based ensemble methods (Random Forest, XGBoost, and CatBoost) is consistent with recent evidence showing their superior performance in DHS-based health prediction tasks due to their ability to model interactions among socioeconomic and demographic variables (Zemariam et al., 2024).

Table 5. Machine-learning model performance

Model	Precision	Recall	F1-score
Logistic Regression	0.83	0.57	0.68
Random Forest	0.82	0.73	0.77
Gradient Boosting	0.80	1.00	0.89
XGBoost	0.83	0.65	0.73
CatBoost	0.83	0.67	0.74
SVM	0.80	1.00	0.89

Determinants of Anaemia Identified Through Machine Learning

Based on the performance of the machine learning model in Table 5, the random forest model was used to identify the relative importance of key predictors contributing to anaemia risk among Nigerian women (Figure 4). Figure 4 shows that body mass index (BMI) is the most influential predictor of anaemia risk, contributing substantially more than any other variable. This finding aligns with evidence from Tirore et al. (2024) and Worku et al. (2023) showing that both undernutrition and overweight-related inflammation are key pathways to anaemia among women in low- and middle-income countries. This reflects the growing "double burden of malnutrition" in sub-Saharan Africa. Also, contraceptive use is an important predictor, which shows the role of reproductive health in anaemia risk (see Figure 4). Hormonal contraceptives are known to reduce menstrual blood loss and thereby lower the likelihood of iron deficiency, whereas non-use may increase vulnerability to anaemia (WHO, 2023). This result supports prior DHS-based studies that identify family-planning uptake as an indirect but significant determinant of women's micronutrient status (Zemariam et al., 2024). Other variables, such as region, residence type, ethnicity, wealth, and religion, emerged as important determinants of anaemia, likely reflecting differences in disease ecology, dietary practices, healthcare access, and social norms across Nigeria (Ogunsakin et al., 2024; Sanni, 2025). However, media exposure, education, household composition, and health insurance showed weaker contributions (Molnar et al., 2024; Khan et al., 2025).

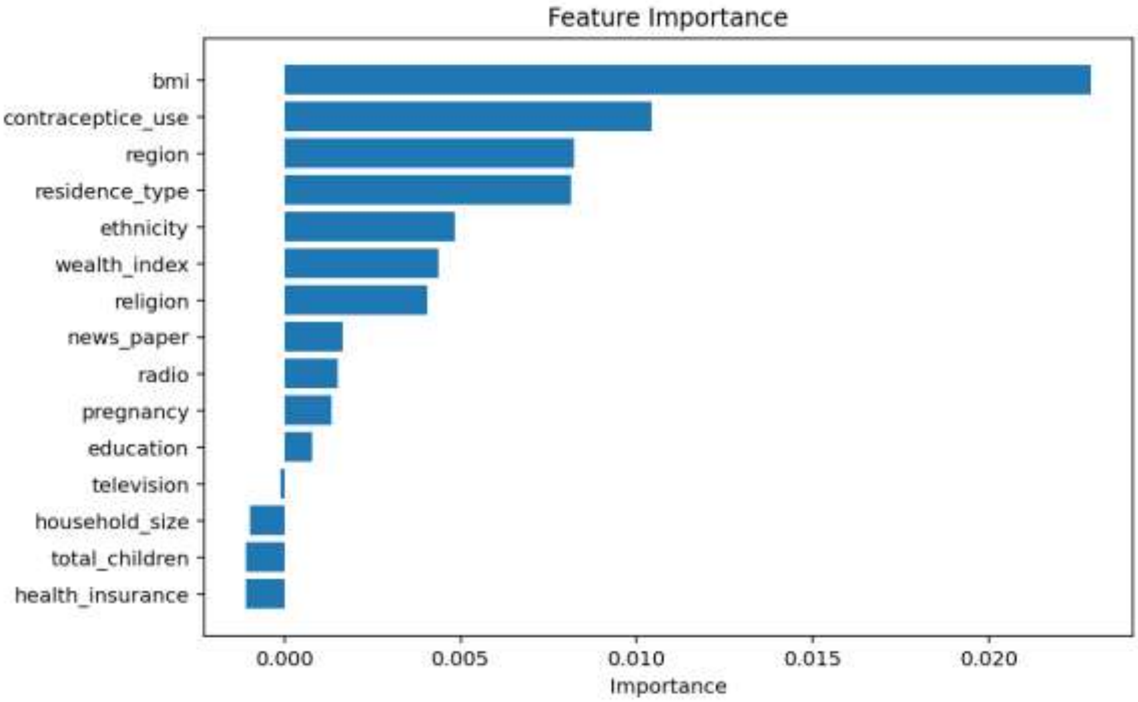

Figure 4. Permutation feature importance ranking of predictors of anaemia among Nigerian women of reproductive age, derived from the machine-learning model

Local Model Explainability Using LIME

Local Interpretable Model-Agnostic Explanations (LIME) was applied to illustrate how specific features contributed to the model's prediction of anaemia for an individual woman, providing insight into case-level decision-making rather than population-level patterns (Figure 3). The most influential contributor is a BMI greater than 0.38, indicating that the individual's nutritional status significantly increases the predicted probability of anaemia. This finding is consistent with evidence in Figure 2. The next most important contributor is pregnancy status, where the condition "Pregnancy ≤ –0.31" (reflecting the individual's standardised value) positively influenced anaemia risk. Pregnancy is a well-established risk factor due to increased iron requirements and haemodilution, especially where supplementation coverage is suboptimal (WHO, 2023). Age also shows a notable contribution, suggesting that the individual's position within the reproductive age range meaningfully shaped risk, even though age did not emerge as a strong discriminator at the population level. This highlights how local explanations can reveal heterogeneous effects that are masked in global analyses (Zafar, 2021).

Contextual and socio-cultural factors, region, religion, and residence type, also contributed positively to the prediction, underscoring that place of residence and cultural context can influence anaemia risk through differences in diet, malaria exposure, sanitation, and healthcare access (Ogunsakin et al., 2021). Contraceptive use appears as a protective–risk modifier, aligning with literature showing that hormonal contraception can reduce menstrual blood loss and lower anaemia risk, while non-use may increase vulnerability (WHO, 2023).

Lower-ranked contributors, such as radio exposure, household size, and total number of children, had smaller but still positive effects. These variables likely act indirectly, reflecting information access, household resource dilution, and cumulative reproductive burden, which have been linked to women's nutritional outcomes in Nigeria (Abubakar et al., 2024).

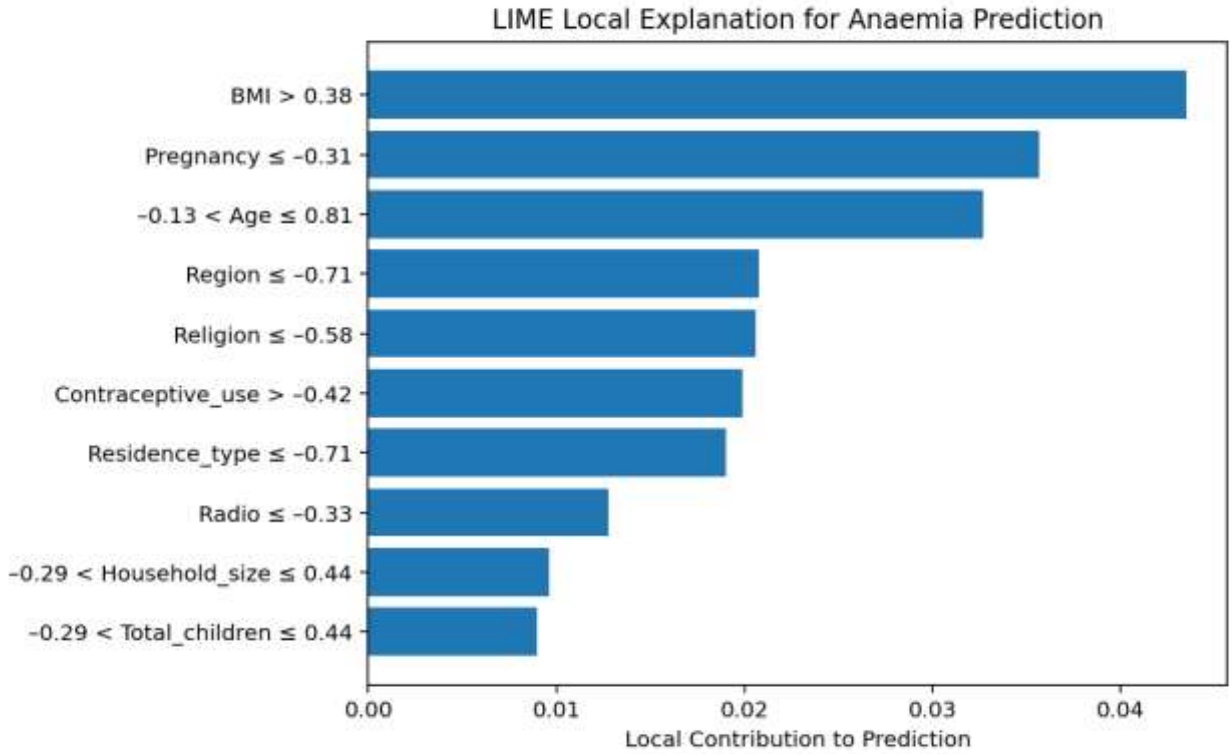

Figure 5. Local Interpretable Model-Agnostic Explanation (LIME)

CONCLUSION AND RECOMMENDATION

In conclusion, anaemia remains a serious public health problem among Nigerian women of reproductive age, with prevalence far exceeding World Health Organisation thresholds. Descriptive and Inferential analysis reflect a high prevalence among wealthier and more educated women, while machine-learning models predicted nutritional status and reproductive factors as the dominant drivers of anaemia risk. Based on these findings, it is recommended that a population-wide, nutrition-sensitive strategies that extend beyond poverty-focused approaches to include food fortification, dietary diversification, and obesity-aware micronutrient interventions is required. Also, strengthening and integrating reproductive health and family-planning services with routine anaemia screening and supplementation is essential. However, future research should prioritise the integration of survey data with biomarkers, infection indicators, and dietary measures to enhance predictive accuracy and support precision public-health interventions.

REFERENCES

1. AbioyeA.A., Owopetu C.A., Adamu-Adedipe F.A., Odesanya L.O., and Olofin-Samuel M.A (2024) Anaemia Prevention Among Pregnant Women, International Journal of Nursing, Midwife and Health Related Cases, 10 (2), 1-11
2. Aboagye, R. G., Okyere, J., Seidu, A. A., Ahinkorah, B. O., Budu, E., & Yaya, S. (2023). Relationship between history of hormonal contraceptive use and anaemia status among women in sub-Saharan Africa: A large population-based study. Plos one, 18(6), e0286392.
3. Abubakar, H. A., Shahril, M. R., & Mat, S. (2024). Nutritional status and dietary intake among Nigerian adolescent: a systematic review. BMC Public Health, 24(1), 1764.
4. Aigbedion, G., Tseng, P. C., & Puthussery, S. (2025). Micronutrient deficiencies and determinants among pregnant women and children in Nigeria: systematic review and meta-analysis. Nutrients, 17(14), 2338.
5. Babah, O. A., Sagastume, D., Akinajo, O. R., Scarpa, G., Hanson, C., Larsson, E. C., ... & Beňová, L. (2025). Dietary diversity insufficiently explains differences in prevalence of anaemia in pregnancy across

regions in Nigeria: A secondary analysis of Demographic and Health Survey 2018. PLOS Global Public Health, 5(5), e0004540.
6. Bello, M., Olarewaju, C. A., & Ayoola, J. I. (2025). Effect of food cultural beliefs on the dietary pattern and food practices of children in Ondo West local area of Ondo state.
7. Çetin, V., & Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 28(2), 299-312.
8. Foster, S. D., Lawrence, E., & Hoskins, A. J. (2024). Spatially Clustered Survey Designs. Journal of Agricultural, Biological and Environmental Statistics, 29(1), 130-146.
9. Gunarathna, S.P., Wickramasinghe, N.D., Agampodi, T.C., Prasanna, I.R. and Agampodi, S.B., 2024. Assessment of the productivity loss due to leading maternal ill-health conditions: a follow-up study of a prospective pregnancy cohort in rural Sri Lanka. BMJ open, 14(10), p.e082798.
10. ICF International (2024) Nigeria Demographic and Health Survey 2023–2024: Key Indicators Report. Rockville, MD: ICF. Available at: https://dhsprogram.com
11. Khan, A., Ali, A., Khan, J., Ullah, F., & Faheem, M. (2025). Using Permutation-Based Feature Importance for Improved Machine Learning Model Performance at Reduced Costs. IEEE Access.
12. Kinyoki, D., Osgood-Zimmerman, A. E., Bhattacharjee, N. V., Kassebaum, N. J., & Hay, S. I. (2021). Anemia prevalence in women of reproductive age in low-and middle-income countries between 2000 and 2018. Nature medicine, 27(10), 1761-1782.
13. Kitaw, B., Asefa, C., & Legese, F. (2024). Leveraging machine learning models for anemia severity detection among pregnant women following ANC: Ethiopian context. BMC public health, 24(1), 3500. https://doi.org/10.1186/s12889-024-21039-x
14. Molnar, Christoph, Gunnar König, Bernd Bischl, and Giuseppe Casalicchio. "Model-agnostic feature importance and effects with dependent features: a conditional subgroup approach." Data Mining and Knowledge Discovery 38, no. 5 (2024): 2903-2941.
15. National Population Commission (NPC) [Nigeria] and ICF (2024) Nigeria Demographic and Health Survey 2023–2024 Final Report. Abuja and Rockville, MD: NPC and ICF.
16. Nigeria Demography and Health Survey, (2024).
17. Obeagu, E. I., & Agreen, F. C. (2023). Anaemia among pregnant women: A review of African pregnant teenagers.
18. Obi, J. C. (2023). A comparative study of several classification metrics and their performances on data. World Journal of Advanced Engineering Technology and Sciences, 8(1), 308-314.
19. Ogbuju, N. E., Anoke, U. R., Olaiya, P. A., Wakama, T., Okoro, L. C., Elebesunu, E. C., ... & Jeremiah, Z. A. (2025). Management of Anaemia during Pregnancy: Insights into Pathophysiology and Public Health Implications in Sub-Saharan Africa. African Journal of Laboratory Haematology and Transfusion Science, 4(1), 58-68.
20. Ogunsakin, R. E., Akinyemi, O., Babalola, B. T., & Adetoro, G. (2021). Spatial pattern and determinants of anemia among women of childbearing age in Nigeria. Spatial and Spatio-temporal Epidemiology, 36, 100396.
21. Ren, Y., Wu, D., Tong, Y., López-DeFede, A., & Gareau, S. (2023). Issue of data imbalance on low birthweight baby outcomes prediction and associated risk factors identification: establishment of benchmarking key machine learning models with data rebalancing strategies. Journal of medical Internet research, 25, e44081.
22. Sanni, S. M. (2025). Nutrition Transition in West Africa: Public Health Challenges, Socioeconomic Impacts, And Regional Policy Interventions. Center of Medicine, 1(1).
23. Shatilwe, J. T., Kuupiel, D., & Mashamba-Thompson, T. P. (2021). Evidence on access to healthcare information by women of reproductive age in low-and middle-income countries: scoping review. Plos one, 16(6), e0251633.
24. Tilahun, W. M., Gebreegziabher, Z. A., Geremew, H., & Simegn, M. B. (2024). Prevalence and factors associated with anemia among HIV-infected women in sub-saharan Africa: a multilevel analysis of 18 countries. BMC Public Health, 24(1), 2236.
25. Tirore, L. L., Areba, A. S., Tamrat, H., Habte, A., & Abame, D. E. (2024). Determinants of severity levels of anemia among pregnant women in Sub-Saharan Africa: multilevel analysis. Frontiers in Global Women's Health, 5, 1367426.

26. Wallace, S. M. (2020). Factors affecting dietary intake, dietary change, nutritional status and appetite in older adults: impact of oral health status (Doctoral dissertation, Queen's University Belfast).
27. Worku, M. G., Alamneh, T. S., Teshale, A. B., Yeshaw, Y., Alem, A. Z., Ayalew, H. G., ... & Tesema, G. A. (2022). Multilevel analysis of determinants of anemia among young women (15-24) in sub-Sahara Africa. Plos one, 17(5), e0268129.
28. World Health Organization (WHO) (2023) Anaemia in Women and Children: Global Estimates and Trends 2000–2022. Geneva: WHO. Available at: https://www.who.int
29. Zafar, M. R., & Khan, N. (2021). Deterministic local interpretable model-agnostic explanations for stable explainability. Machine Learning and Knowledge Extraction, 3(3), 525-541.
30. Zemariam, A. B., Yimer, A., Abebe, G. K., Wondie, W. T., Abate, B. B., Alamaw, A. W., ... & Ngusie, H. S. (2024). Employing supervised machine learning algorithms for classification and prediction of anemia among youth girls in Ethiopia. Scientific reports, 14(1), 9080.