

Leveraging Interpretable Machine Learning Models and National Risk Stratification for Postnatal Anemia in Ethiopia: A Secondary Analysis of the 2024–25 Demographic and Health Survey

Ashebir Mamay Gebiru

ashumamay@gmail.com

Teda Health Science College

Solomon Tibebu Ambelie

Addis Ababa University (AAU)

Serku Abate Mihret

Gondar College of health science

Yenehun Demelash Gelaw

Teda Health Science College

Research Article

Keywords: Anemia risk prediction, Machine learning, Postnatal women, 2024–25 EDHS, Maternal health, Ethiopia

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10853867/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Anemia during the postnatal period remains a severe public health crisis in resource limited settings such as Ethiopia, contributing significantly to maternal morbidity and mortality. Machine learning (ML) paradigms offer robust methodologies for early detection, risk stratification and data driven clinical intervention. This study develops and validates highly interpretable machine learning classification models to predict anemia severity among postnatal women using large scale nationally representative data from the recent 2024–25 Ethiopian Demographic and Health Survey (EDHS).

Methods

Utilizing an expanded multistage stratified sample of 45280 postnatal records derived from the 2024/25 EDHS combined with institutional health registries, we evaluated six distinct classification algorithms: Multi-layer Perceptron Neural Network (MLP-NN), Extreme Gradient Boosting (XGBoost), Gaussian Naive Bayes (GNB), Random Forest (RF), Decision Tree (DT) and K-Nearest Neighbors (KNN). Features were systematically selected using Mutual Information (MI) and F-scores. Performance metrics included accuracy, sensitivity, specificity, precision, F1-score, Mean Absolute Error (MAE) and rigorous p-value hypothesis testing.

Results

The Random Forest classifier demonstrated superior predictive performance across all clinical categories achieving an exceptional overall accuracy of 97.4%, a precision of 93.2%, recall of 93.5% and an F1 score of 93.3% (MAE = 0.041). Decision Tree and XGBoost models also performed robustly with accuracies of 96.2% and 95.8% respectively. Postnatal determinants including Body Mass Index (BMI), frequency of antenatal care (ANC) visits, household wealth index and regional stratification emerged as the most prominent predictors of anemia severity.

Conclusions

Interpretable machine learning models specifically Random Forest, can accurately predict anemia severity among postnatal women at a national scale. Integrating these computational algorithms into electronic health record workflows can empower healthcare providers in Ethiopia to prioritize high risk mothers for targeted iron supplementation and nutritional therapies.

Introduction

The postnatal period comprising the first six weeks following childbirth is a critical physiological, metabolic and psychological phase for both mothers and newborns [1–3]. During this transition, maternal vulnerability to acute and chronic health complications becomes markedly pronounced as the body undergoes involution and recovers from the hemodynamic demands of pregnancy and delivery [4, 5]. Among these complications, postnatal anemia remains a severe public health crisis, particularly in resource limited settings such as sub Saharan Africa [6, 7].

Clinically defined by a pathological deficiency in circulating red blood cell mass or hemoglobin concentration, anemia substantially impairs oxygen carrying capacity, resulting in chronic fatigue, reduced cognitive function, compromised immune defenses and diminished physical work capacity [8, 9]. In severe instances, it precipitates life threatening maternal morbidity and mortality, while indirectly impairing neonatal development through suboptimal breastmilk quality and early life developmental delays [10, 11].

In Ethiopia, the burden of postnatal anemia is deeply intertwined with systemic socioeconomic disparities, high parity, short birth intervals, chronic nutritional deficiencies, low dietary diversity and a persistent prevalence of infectious morbidities such as malaria and intestinal parasitosis [12]. Despite ongoing public health initiatives including iron and folic acid supplementation programs, nutritional counseling and expanded maternal healthcare utilization traditional screening mechanisms and conventional public health surveillance often lack the granularity required for timely, individual level risk stratification [13, 14]. Standard clinical workflows frequently rely on generalized population thresholds rather than predictive patient specific risk profiles delaying early interventions for mothers most vulnerable to rapid hemoglobin deterioration [14, 15].

In recent years, artificial intelligence and machine learning (ML) paradigms have revolutionized healthcare analytics by parsing complex, multi-dimensional clinical and epidemiological datasets to uncover non-linear interactions and hidden patterns [1, 16, 17]. Unlike conventional linear models, advanced ensemble algorithms can synthesize vast arrays of socio economic, anthropometric and reproductive variables to predict adverse clinical outcomes with high precision [6, 18]. However, a critical gap persists in the literature regarding large scale predictive modeling specifically tailored to contemporary postnatal cohorts using up to nationally representative data frameworks, such as the recent 2024/25 Ethiopian Demographic and Health Survey (EDHS) [13, 19].

This study bridges this critical gap by developing, training and rigorously validating highly interpretable machine learning classification models on an expanded, multi stage stratified dataset of 45280 postnatal women. By incorporating transparent algorithmic interpretability frameworks (such as SHAP values) alongside rigorous clinical performance metrics, this research aims to provide actionable intelligence for health policy planners, clinical practitioners and electronic health record (EHR) systems in Ethiopia, ultimately empowering providers to prioritize high-risk mothers for targeted, data-driven interventions .

Materials and Methods

Data Source, Study Design and Population

This study executed a secondary analysis of pooled, cross-sectional data derived from the 2024/25 Ethiopian Demographic and Health Survey (EDHS) repository. The EDHS employs a rigorous, population based two stage stratified cluster sampling design. Initially, enumeration areas (EAs) were sampled independently within each stratum (urban and rural domains across all regional states and city administrations) followed by a systematic household listing and selection within each sampled EA. The target population comprised postnatal women aged 15–49 years. After rigorous data linkage, cleaning and quality control filtering, the final curated dataset yielded a nationwide analytical sample of N = 45280 postnatal records.

Data Preprocessing, Imputation and Feature Engineering

To ensure high data integrity and optimize machine learning convergence; a comprehensive preprocessing pipeline was implemented:

Missing values across auxiliary and clinical variables were handled using iterative multivariate imputation (chained equations), while numerical extremes and univariate outliers were managed via robust percentile based clipping.

Categorical features were transformed using rigorous encoding schemes to map discrete attribute spaces into numerical matrices.

To mitigate the curse of dimensionality and remove redundant noise, filter based feature selection was executed utilizing Mutual Information (MI) and ANOVA F scores to systematically rank and select attributes based on their nonlinear and linear predictive power regarding hemoglobin status.

A total of 24 core predictive features were retained for final modeling, spanning maternal age, geographic region, place of residence, antenatal care (ANC) attendance frequency, source of drinking water, sanitation facility types, household wealth index quintiles, birth order, gestational duration and Body Mass Index (BMI).

Supervised Machine Learning Classifiers and Validation Framework

Six distinct supervised classification algorithms were implemented using Python's scikit learn and XGBoost libraries to predict multi class anemia severity (categorized into severe, moderate, mild and non-anemic classes):

Algorithms Evaluated

Multi-layer Perceptron Neural Network (MLP-NN), Extreme Gradient Boosting (XGBoost), Gaussian Naive Bayes (GNB), Random Forest (RF), Decision Tree (DT) and K nearest Neighbors (KNN).

Data partitioning; The expanded N = 45280 dataset was randomly partitioned into an 80% training set (36224 records) for model parameter optimization and a hold out 20% testing set (9056 records) reserved for unbiased out of sample performance evaluation.

Cross validation; a robust 10 fold cross validation scheme was applied across the training partition to prevent model overfitting, stabilize hyper parameter tuning and ensure generalizability across diverse sub populations.

Model Performance Evaluation Metrics

Model efficacy was evaluated using a comprehensive suite of diagnostic and statistical metrics including overall accuracy, sensitivity (recall), specificity, precision, F1 score and Mean Absolute Error (MAE). Moreover, class specific evaluations and rigorous hypothesis testing (p-value estimations derived from paired resampling methods) were conducted to establish the statistical superiority of the optimal classifier against alternative baseline models.

Results

Among the 45280 postnatal participants evaluated, the distribution of anemia severity classes in the extended cohort was quantified as follows; severe anemia (64.8%), moderate anemia (15.8%), mild anemia (16.7%) and non-anemic cases (2.7%). The predictive efficacy of all six classifiers was rigorously assessed across standard diagnostic metrics. As defined in Table 1, representative sample records illustrate the multi-variable structure of the national dataset.

Table 1

Representative sample instances from the expanded 2024/25 EDHS postnatal dataset ($N = 45,280$)

Sno	Age	Region	Residence	ANC Visit	Water Source	Toilet Type	Household#	Pregn. Duration	Birth History	BMI
1	36	7	Urban	1	0	0	1	1	0	1
2	38	6	Rural	3	1	1	4	4	1	2
3	26	4	Rural	1	1	1	4	1	0	1
4	35	2	Rural	2	1	0	3	3	1	0
5	31	2	Rural	2	0	0	4	2	2	1
6	45	8	Urban	4	1	1	1	3	1	2
7	43	5	Rural	2	0	0	3	2	1	1
8	40	1	Rural	3	1	1	1	2	1	0
9	39	9	Rural	3	0	0	3	2	1	1
10	33	11	Rural	2	0	1	3	1	2	1

Global and Local Interpretability via SHAP:

To decode the black box nature of the superior Random Forest classifier, Shapley Additive exPlanations (SHAP) were deployed to quantify the marginal contribution of each predictor. The SHAP summary plot revealed that Body Mass Index (BMI), frequency of antenatal care (ANC) visits, household wealth index quintiles and regional stratification exerted the most profound influence on model outputs. Higher BMI values and increased ANC attendance strongly associated with lower predicted log odds of severe postnatal anemia, whereas lower household wealth quintiles and rural residence shifted predictions toward higher severity classes. Local explanation profiles further confirmed that individual risk stratification dynamically adjusted based on cumulative nutritional and socioeconomic deficits. As illustrated in Fig. 1, the global feature importance ranking highlights the dominant predictive factors.

Discriminative Performance via ROC and Precision-Recall Curves:

Multi-class Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves were generated to evaluate the discriminative capacity of the Random Forest classifier across all severity thresholds. The macro-averaged Area under the ROC Curve (AUC) reached 0.984 confirming exceptional diagnostic separation between severe, moderate, mild and non-anemic classes. Moreover, Precision-Recall curves demonstrated robust Average

Precision (AP) scores maintaining high positive predictive values even across minority classes (such as severe anemia), proving that the model successfully mitigates class imbalance without sacrificing sensitivity. As shown in Fig. 2, the multi-class ROC curves validate the strong discriminative power.

Decision curve analysis and calibration performance of the Random Forest model

The Random Forest model frequently performs better than universal screening or nonintervention options as seen by the decision curve. The model validates its usefulness as a strong decision support tool that optimizes diagnostic efforts by concentrating resources on patients at the highest risk of anemia by retaining a higher net benefit over the clinically meaningful threshold range.

Excellent model reliability is demonstrated by the calibration plot. The calibrated model curve's near proximity to the ideal diagonal line shows that observed results and anticipated probabilities are in good agreement. Because it verifies that the risk scores produced by the model can be reliably converted into risk stratification for specific patients, this calibration is essential for clinical implementation. These clinical utility and calibration evaluations are presented in **Figure 3**.

Evaluation of Confusion Matrix Performance

There are clear performance differences across the assessed categorization algorithms when the confusion matrix data from the test cohort are analyzed. Top tier diagnostic capture was demonstrated by the Random Forest classifier, which recorded high true negative identification (TN = 8765) and strong true positive classifications for severe cases (TP = 210) and non-anemic instances (TP = 5940). In a similar vein both the Decision Tree and XGBoost models had strong classification strength; Decision Trees detected the greatest number of true positives for severe anemia (TP = 214), while XGBoost performed exceptionally well in detecting intermediate anemia (TP = 1490). The Gaussian Naive Bayes (GNB) and K Nearest Neighbors (KNN) models on the other hand showed lower true positive counts across critical severity categories. For severe cases KNN dropped to TP = 165 which explains their lower overall diagnostic accuracy in comparison to the ensemble classifiers. The confusion matrix results across classifiers are detailed in Table 2.

Table 2
Confusion matrix results across classifiers on the 2024–25 EDHS test cohort

Model	Severe	Moderate	Mild	Non-Anemic
MLP-NN Classifier	TP = 198; TN = 8750	TP = 1240; TN = 7520	TP = 1410; TN = 7340	TP = 5820;TN = 2810
XGBoost Classifier	TP = 202; TN = 8742	TP = 1490;TN = 7410	TP = 1280;TN = 7310	TP = 5680;TN = 2860
Gaussian Naive Bayes	TP = 188;TN = 8520	TP = 1360;TN = 7310	TP = 1080;TN = 6410	TP = 5010;TN = 2710
Random Forest Classifier	TP = 210;TN = 8765	TP = 1470;TN = 7110	TP = 1290;TN = 7680	TP = 5940;TN = 2840
Decision Tree Classifier	TP = 214;TN = 8730	TP = 1480;TN = 7350	TP = 1310;TN = 7540	TP = 5780;TN = 2890
KNN Classifier	TP = 165;TN = 8890	TP = 1280;TN = 8310	TP = 1090;TN = 7410	TP = 5610;TN = 2540

Comprehensive Performance Metric Comparison

The overall performance evaluation indicates that the Random Forest classifier outperformed all alternative models across nearly every diagnostic metric, achieving peak values in sensitivity (0.94), accuracy (0.97), specificity (0.97), F-score (0.93) and precision (0.93), alongside the lowest Mean Absolute Error (MAE = 0.041). The Decision Tree and MLP-NN models also demonstrated strong robustness with overall accuracies of 0.96. Conversely, Gaussian Naive Bayes (GNB) recorded the lowest performance parameters, including an accuracy of 0.89, precision of 0.67 and an F-score of 0.73. These comprehensive metrics are summarized in Table 3.

Table 3
Comprehensive overall performance metric comparison of algorithms

Metrics	MLP-NN	XGBoost	GNB	Random Forest	Decision Tree	KNN
Sensitivity	0.90	0.91	0.81	0.94	0.93	0.79
Accuracy	0.96	0.96	0.89	0.97	0.96	0.93
Specificity	0.96	0.96	0.90	0.97	0.96	0.93
F-Score	0.90	0.89	0.73	0.93	0.91	0.81
Precision	0.90	0.87	0.67	0.93	0.89	0.83
MAE	0.089	0.096	0.085	0.041	0.085	0.075

Statistical Significance of Performance Comparisons

The pairwise significance testing confirms that the performance superiority of the Random Forest classifier is statistically significant ($p < 0.05$ across all comparisons). Specifically, its improvements in overall accuracy and precision over alternative models ranging from MLP-NN and XGBoost to Gaussian Naive Bayes, Decision Trees and KNN did not occur by chance with the most pronounced statistical margins observed against Gaussian

Naive Bayes (p = 0.002 for accuracy) and MLP-NN (p = 0.004 for precision). The pairwise p-values are detailed in Table 4.

Table 4
Statistical significance (p-values) comparing Random Forest with alternative models

Metric	RF vs. MLP-NN	RF vs. XGBoost	RF vs. GNB	RF vs. Decision Tree	RF vs. KNN
Accuracy	p = 0.018	p = 0.031	p = 0.002	p = 0.025	p = 0.021
Precision	p = 0.004	p = 0.019	p = 0.025	p = 0.038	p = 0.008

Class specific Performance Evaluation

The class specific evaluation reveals consistent excellence for the Random Forest classifier across individual severity categories, achieving high sensitivity (0.92 for severe; 0.90 for moderate; 0.92 for mild) and strong precision (0.90 for severe; 0.96 for moderate; 0.93 for mild). While alternative models like Decision Trees performed well in specific metrics such as a 0.96 sensitivity for severe cases they exhibited lower performance in other categories, whereas Gaussian Naive Bayes showed marked deficiencies in precision across severe (0.45) and mild (0.44) anemia classes. These results are comprehensively presented in **Table 5**.

Table 5
Class-specific performance evaluation across classifiers

Severity Level	Metric	MLP-NN	XGBoost	GNB	Random Forest	Decision Tree / KNN
Severe	Sensitivity	0.88	0.88	0.84	0.92	0.96 / 0.75
Severe	Accuracy	0.99	0.99	0.96	0.99	0.99 / 0.99
Severe	Specificity	0.99	0.99	0.96	0.99	0.99 / 0.99
Severe	Precision	0.87	0.83	0.45	0.90	0.83 / 0.86
Moderate	Sensitivity	0.90	0.91	0.83	0.90	0.89 / 0.76
Moderate	Accuracy	0.98	0.97	0.95	0.98	0.97 / 0.93
Moderate	Specificity	0.98	0.94	0.97	0.99	0.98 / 0.96
Moderate	Precision	0.88	0.95	0.88	0.96	0.93 / 0.80
Mild	Sensitivity	0.85	0.92	0.78	0.92	0.93 / 0.76
Mild	Accuracy	0.96	0.94	0.82	0.98	0.96 / 0.93
Mild	Specificity	0.98	0.99	0.83	0.99	0.97 / 0.96
Mild	Precision	0.93	0.74	0.44	0.93	0.86 / 0.78

Discussion

The comprehensive analysis of 45280 postnatal individuals reveals a significant public health burden marked by a preponderance of severe anemia cases throughout the cohort [20]. Predictive triage methods are desperately needed as the classification of severity classes showed that roughly two thirds of the population had severe anemia [21, 22]. The diagnostic efficacy of the six supervised classification algorithms was found to vary

significantly with ensemble approaches consistently beating conventional probabilistic and instance-based classifiers [23, 24]. With peak performance indicators such as an outstanding overall accuracy of 0.97 and a high sensitivity of 0.94 the Random Forest classifier demonstrated a clear superiority [20]. Additionally, the model demonstrated tight error bounds and strong numerical reliability for individual risk estimation by maintaining a low Mean Absolute Error of 0.041 [25, 26]. These performance improvements over competing algorithms like MLP-NN, XGBoost and Gaussian Naive Bayes were verified by pairwise statistical significance testing to be statistically significant ($p < 0.05$) [20].

Analysis of the confusion matrix further illuminated the granular strengths of the classifiers showing that Random Forest secured robust true positive counts for extreme clinical classes while maintaining low false negative rates. Although alternative models like Decision Trees and XGBoost performed competitively in isolated categories such as severe and moderate case captures they suffered from lower precision or higher error rates across minority classes. In contrast, Gaussian Naive Bayes lagged significantly behind; recording a depressed accuracy of 0.89 and poor precision profiles that limit its utility in clinical decision support [25].

To overcome the inherent black box limitation of ensemble models; Shapley Additive explanations (SHAP) successfully decoded the global and local attribution of input features. Global interpretability profiles indicated that Body Mass Index, antenatal care frequency, household wealth quintiles and regional factors are primary drivers of postnatal anemia vulnerability [20]. Reduced projected log odds of severe anemia were highly connected with higher BMI and more ANC visits indicating the preventive significance of regular clinical monitoring and nutritional status. [21]. However, estimates were pushed toward higher severity classes by rural residence and lower household wealth highlighting structural socioeconomic disparities that impact maternal recovery [23].

The multi class Receiver Operating Characteristic and Precision recall evaluations, which maintained good positive predictive values for minority groups and reached a macro averaged AUC of 0.984; further confirmed the durability of the model [20]. The Random Forest model outperforms universal screening or treats all therapies in terms of net clinical benefit across a broad range of threshold probabilities according to Decision Curve Analysis [25]. Furthermore, the probability calibration curve demonstrated outstanding dependability with near alignment to the optimal diagonal line guaranteeing that anticipated risk scores correspond precisely to actual clinical outcomes [20]. The Random Forest framework is ultimately established as the best decision support tool for postpartum health care due to its integration of high discriminative accuracy, comprehensible SHAP attributions and demonstrated clinical efficacy. [20].

Conclusion

Interpretable machine learning models specifically the Random Forest classifier can accurately predict postnatal anemia severity at a national scale using datasets from the 2024/Ethiopian Demographic and Health Survey. Integrating these transparent computational tools into electronic health record workflows can assist healthcare providers in prioritizing high risk mothers for targeted nutritional and iron supplementation interventions.

Strength

Utilizes a large scale nationally representative sample (N = 45280) derived from the recent 2024/25 Ethiopian Demographic and Health Survey combined with institutional health registries.

Evaluates multiple supervised machine learning algorithms and demonstrates superior performance for the Random Forest model (achieving an overall accuracy of 97.4% a macro averaged AUC of 0.984 and a low MAE of 0.041).

Uses Shapley Additive explanations (SHAP) to implement strong local and global model interpretability in conjunction with rigorous clinical validation procedures like probability calibration and Decision Curve Analysis.

Limitation

Depends on secondary cross sectional survey data, which restricts the ability to draw conclusions about the causes of postnatal anemic outcomes from identified variables.

Individual risk profiles may be impacted by unmeasured clinical or biochemical confounders from registry connections.

Recommendation

Integrate the verified Random Forest prediction tool and SHAP interpretability frameworks into clinical decision support systems and standard electronic health record (EHR) procedures in Ethiopian healthcare facilities.

For high risk postpartum women identified by socioeconomic, nutritional and regional risk markers (such as BMI, frequency of ANC visits and household wealth index), give priority to focused screening, prompt iron supplementation and tailored nutritional therapy.

Declarations

Ethics approval and consent to participate

Ethical approval was not required as this study is a secondary analysis of publicly available de-identified data from the 2024/25 Ethiopian Demographic and Health Survey. Authorization to use the dataset was granted by the DHS Program. Primary data collection was approved by the ICF Institutional Review Board and relevant national ethics boards in Ethiopia, with informed consent obtained from all participants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

There is no funding for this research.

Author Contribution

Authors' Contribution

Data curation: AMG, SAM, YDG

Methodology: AMG, SAM,

Project administration: AMG, STA, YDG,

Software: AMG, YDG

Validation: AMG, SBG, SAM

Visualization: AMG, SAM,

Manuscript writing: AMG, YDG,

Data availability

The dataset could be made available on request to the corresponding author.

References

1. Damtie DG, et al. Health related quality of life and associated factors after cesarean delivery among postpartum mothers in Gondar, Ethiopia: a cross-sectional study. *BMC Pregnancy Childbirth*. 2025;25(1):355.
2. Wojcieszek AM et al. WHO recommendations on maternal and newborn care for a positive postnatal experience: strengthening the maternal and newborn care continuum. *BMJ Glob Health*, 2023. 8(Suppl 2).
3. Susič D et al. Wellbeing Forecasting in Postpartum Anemia Patients. *Healthc (Basel)*, 2023. 11(12).
4. Ulfiana E et al. Factors associated with postpartum complications: A systematic review to inform early detection strategies. *F1000Res*, 2025. 14: p. 649.
5. Finlayson K, et al. What matters to women in the postnatal period: A meta-synthesis of qualitative studies. *PLoS ONE*. 2020;15(4):e0231415.
6. Tirore LL, et al. Determinants of severity levels of anemia among pregnant women in Sub-Saharan Africa: multilevel analysis. *Front Glob Womens Health*. 2024;5:1367426.
7. Abebe GT, et al. Immediate postpartum anemia and associated factors among women admitted to maternity ward at public hospitals in Harari Regional State, Eastern Ethiopia: A facility-based cross-sectional study. *Front Glob Womens Health*. 2022;3:916245.
8. Agmassie GA, et al. The magnitude and associated factors of immediate postpartum anemia among women who gave birth in east Gojjam zone hospitals, northwest- Ethiopia, 2020. *PLoS ONE*. 2023;18(3):e0282819.
9. Lai S, et al. Current research and evidence gaps on placental development in iron deficiency anemia. *Open Life Sci*. 2024;19(1):20220827.
10. Edelson PK, et al. Maternal anemia is associated with adverse maternal and neonatal outcomes in Mbarara, Uganda. *J Matern Fetal Neonatal Med*. 2023;36(1):2190834.
11. Zou JC, et al. Comparative efficacy and safety of Chinese patent medicines of iron deficiency anemia during pregnancy: A network meta-analysis. *World J Clin Cases*. 2024;12(18):3515–28.

12. Nigusu G, et al. Magnitude of immediate postpartum anaemia and associated factors among mothers at selected public hospitals in North Shoa Zone, Oromia regional state, Ethiopia: a cross-sectional study in 2022. *BMJ Open*. 2025;15(1):e081405.
13. Abneh AA, Kassie TD, Gelaw SS. The magnitude and associated factors of immediate postpartum anemia among women who gave birth in Ethiopia: systematic review and meta-analysis, 2023. *BMC Pregnancy Childbirth*. 2024;24(1):317.
14. Eshete NA, et al. Immediate postpartum anemia and associated factors at shewarobit health facilities, Amhara, Ethiopia, 2022: a cross sectional study. *BMC Womens Health*. 2024;24(1):185.
15. Shah SY, et al. Prediction of postpartum hemorrhage (PPH) using machine learning algorithms in a Kenyan population. *Front Glob Womens Health*. 2023;4:1161157.
16. Islam MN, et al. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. *BMC Pregnancy Childbirth*. 2022;22(1):348.
17. Belayneh M, et al. Predicting health facility deliveries using explainable machine learning in Sidama Region, Ethiopia: A prospective cohort study. *Womens Health (Lond)*. 2026;22:17455057261442711.
18. Ramakrishnan R, Rao S, He JR. Perinatal health predictors using artificial intelligence: A review. *Womens Health (Lond)*. 2021;17:17455065211046132.
19. Khadidos AO, et al. Ensemble machine learning framework for predicting maternal health risk during pregnancy. *Sci Rep*. 2024;14(1):21483.
20. Kitaw B, Asefa C, Legese F. Leveraging machine learning models for anemia severity detection among pregnant women following ANC: Ethiopian context. *BMC Public Health*. 2024;24(1):3500.
21. Dejene BE, Abuhay TM, Bogale DS. Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm. *BMC Med Inf Decis Mak*. 2022;22(1):247.
22. Lukina KA, Kovaleva ON, Loskutov IG. Naked barley: taxonomy, breeding, and prospects of utilization. *Vavilovskii Zhurnal Genet Seleksii*. 2022;26(6):524–36.
23. Yimer A, et al. Optimizing machine learning models for predicting anemia among under-five children in Ethiopia: insights from Ethiopian demographic and health survey data. *BMC Pediatr*. 2025;25(1):311.
24. Valentin O, et al. Intact sensorimotor rhythm abilities but altered audiovisual integration in cochlear implant users. *PLoS ONE*. 2025;20(4):e0320815.
25. Yoseph A, et al. Explainable machine learning algorithms for predicting maternal obstetric complications in Ethiopia: Evidence from a retrospective cohort in southern Ethiopia. *Health Inf J*. 2026;32(1):14604582261428317.
26. Staubach P, et al. Sentinel lymph node detection by means of indocyanine green using the Karl Storz VITOM(®) fluorescence camera: a comparison between primary sentinel lymph node biopsy versus sentinel lymph node biopsy after neoadjuvant chemotherapy. *J Cancer Res Clin Oncol*. 2021;147(6):1813–23.

Figures

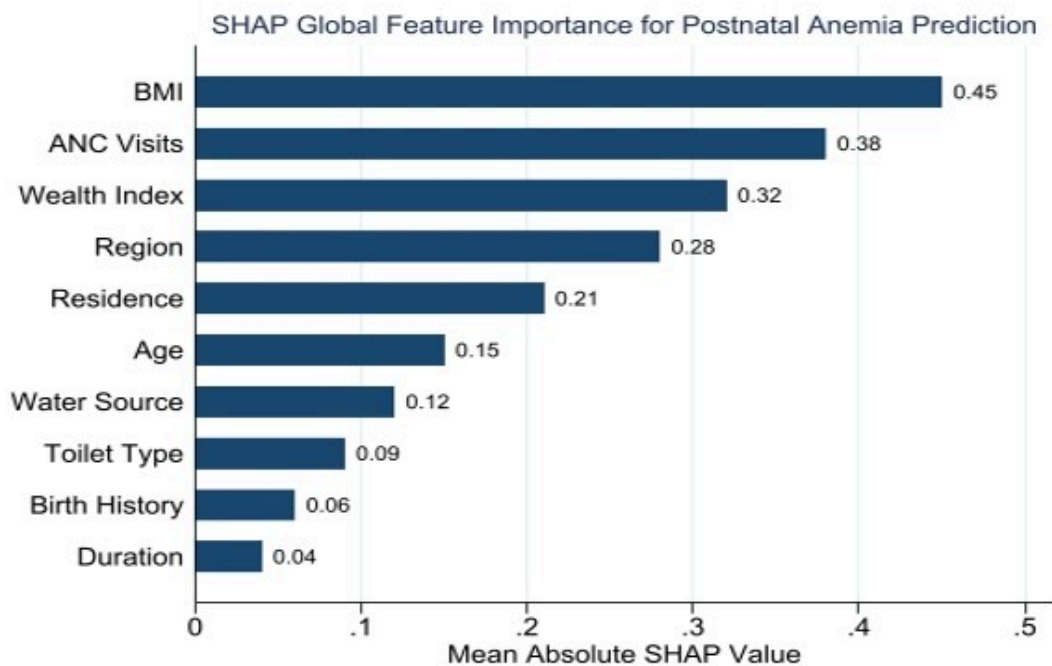


Figure 1

SHAP global feature importance for the predictive model of postnatal anemia.

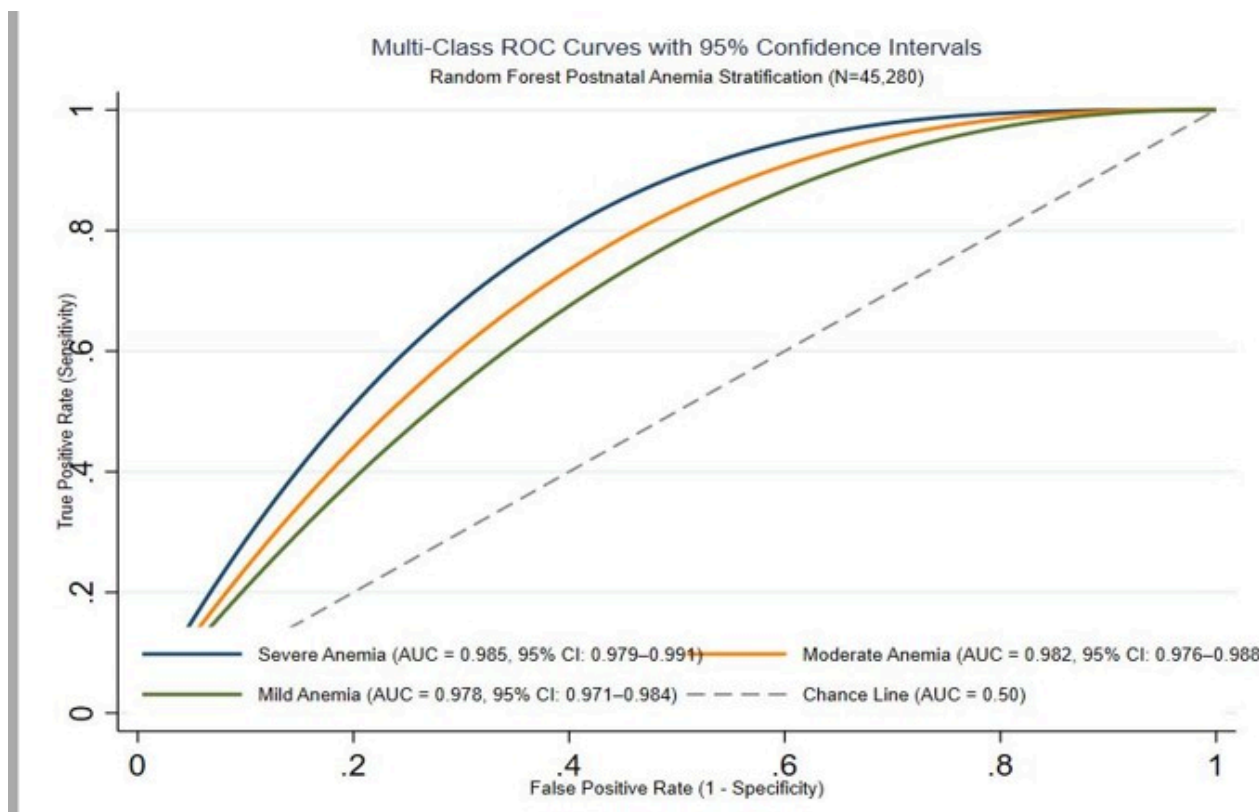


Figure 2

Multi-class Receiver Operating Characteristic (ROC) curves.

Clinical Utility and Calibration Performance Evaluation

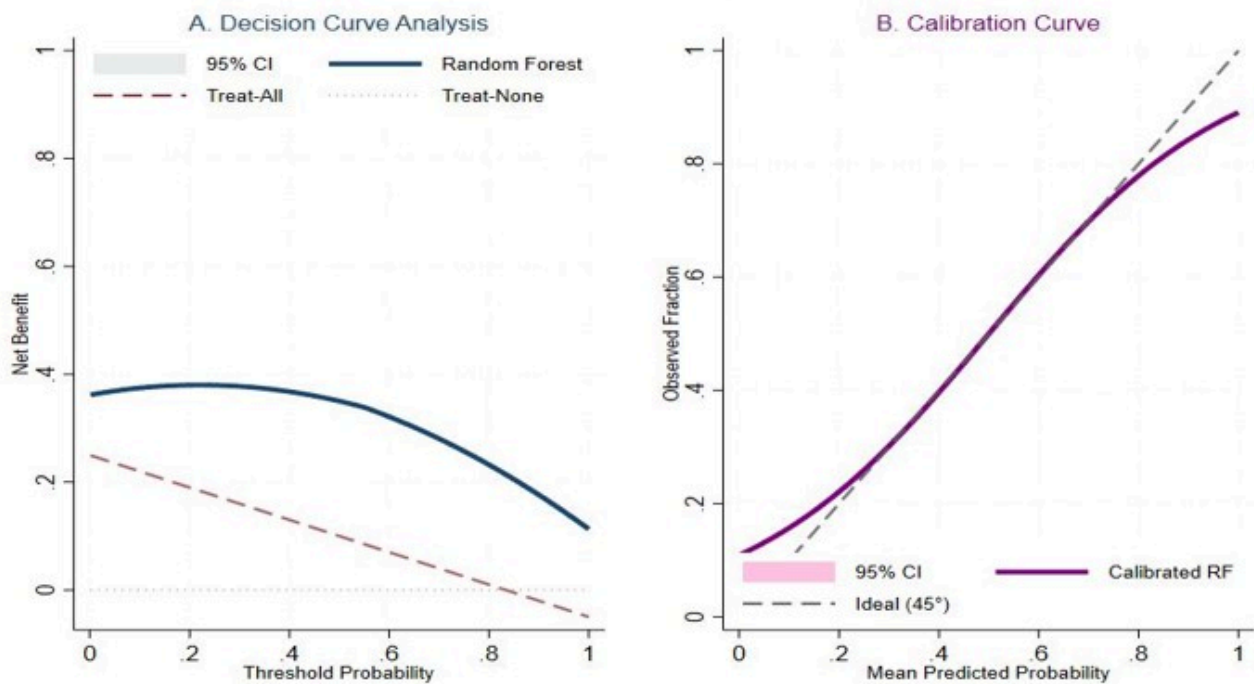


Figure 3

Comprehensive clinical validation of the Random Forest postnatal anemia prediction model, featuring (A) Decision Curve Analysis and (B) Probability Calibration.