Indonesian Journal of Global Health Research
Volume 8 Number 6, December 2026
e-ISSN 2715-1972; p-ISSN 2714-9749
http://jurnal.globalhealthsciencegroup.com/index.php/IJGHR

PREMATURITY PREDICTION MODEL BASED ON NUTRITIONAL STATUS AND ANAEMIA AMONG PREGNANT WOMEN

Dea Zara Avila[1], Nur Husnul Khatimah[1]*, Resti Dona Saputri[2]
[1]Universitas Muhammadiyah Bima, Jl. Anggrek, Nae, Rasanae Barat, Kab. Bima, Nusa Tenggara Barat 84111, Indonesia
[2]STIKES Yahya Bima, Sambinae, Mpunda, Kabupaten Bima, Nusa Tenggara Barat 84111, Indonesia
*nurhusnul62@gmail.com

ABSTRACT
Preterm birth is a maternal-neonatal health problem that contributes to morbidity, mortality, and long-term healthcare needs. This study aimed to develop and conduct preliminary internal validation of a preterm birth prediction model based on chronic energy deficiency (KEK) status and anaemia among pregnant women in Bima City. This quantitative analytical study developed a prediction model for preterm birth using logistic regression. A total of 44 pregnant women were recruited through purposive sampling, based on predetermined inclusion criteria and the availability of complete data on birth outcomes, chronic energy deficiency (CED/KEK) status, anaemia, maternal age, parity, and antenatal care (ANC) visits. The analyses comprised descriptive and bivariate analyses, multivariable logistic regression, discrimination assessment using the receiver operating characteristic-area under the curve (ROC-AUC), calibration assessment using the Brier Score, threshold determination using the Youden Index, and internal validation through bootstrap resampling. Five of 44 respondents (11,36%) experienced preterm birth. The main model yielded the equation Logit(P) = -2,9517 + 0,2041(KEK) + 3,3171(Anemia). KEK yielded an adjusted odds ratio (AOR) of 1,226 (95% CI 0,069-21,715; p=0,889), while anaemia yielded an AOR of 27,579 (95% CI 2,800-271,677; p=0,004). The apparent ROC-AUC of 0,7487 decreased to 0,6725 after optimism correction using bootstrap. The apparent Brier Score was 0,0697. At a threshold of 0,5903, model sensitivity was 60,0% and specificity was 94,9%. The combination of KEK and anaemia can be formulated into a probabilistic model for predicting preterm birth, with anaemia providing the strongest predictive signal in the study sample. However, the limited number of preterm birth events resulted in unstable estimates and performance optimism. The model should be positioned as a pilot model and requires further development in a larger sample and external validation before being used for risk stratification in ANC services.

Keywords: anaemia; chronic energy deficiency; logistic regression; prediction model; preterm birth

How to cite (in APA style)
Avila, D. Z., Khatimah, N. H., & Saputri, R. D. (2026). Prematurity Prediction Model Based on Nutritional Status and Anaemia among Pregnant Women. *Indonesian Journal of Global Health Research*, *8*(6), 291–302. https://doi.org/10.37287/ijghr.v8i6.2628.

INTRODUCTION

Preterm birth is one of the major challenges in maternal and neonatal health. Preterm birth is generally defined as birth before 37 completed weeks of gestation (Fente et al., 2023; Kassahun et al., 2024). Globally, prematurity continues to contribute substantially to neonatal mortality and various long-term health consequences among surviving infants, making the identification of maternal risk factors during pregnancy an important component of prevention efforts (Chun et al., 2025; Kassahun et al., 2024).

One maternal factor of concern is anaemia during pregnancy. Anaemia during pregnancy is associated with various adverse maternal and neonatal outcomes, including preterm birth and low birth weight (Beressa et al., 2024; Khan et al., 2024; Khezri et al., 2023; Wang et al., 2025; Q.-Y. Yu et al., 2024; X. Yu et al., 2026). Biologically, this relationship can be understood through reduced oxygen-carrying capacity, impaired maternal-placental oxygenation, and vulnerability resulting from the nutritional deficiencies underlying anaemia. In addition to anaemia, maternal nutritional status is an important determinant of pregnancy health. Nutritional deficiencies may affect maternal physiological adaptation, placental growth, and nutrient supply to the fetus (González-Fernández et al., 2024). Pre-pregnancy weight status has also been reported to be

associated with prematurity and other pregnancy outcomes (Jiang et al., 2025; Sawadogo et al., 2024). In the Indonesian context, maternal anthropometric indicators such as mid-upper arm circumference (MUAC/LILA) are also relevant for describing nutritional reserves and maternal vulnerability (Adigama et al., 2025).

Most previous studies have focused on estimating associations between maternal factors and pregnancy outcomes. From a public health perspective, information on associations cannot always be directly translated into individual-level decisions. Prediction modelling allows various maternal characteristics to be combined into an individual outcome probability. Research in Indonesia has also shown that pregnancy characteristics can be integrated into the development of maternal health prediction models (Nadhiroh et al., 2024). However, prediction model development must account for sample size, number of outcome events, overfitting, discrimination, calibration, and internal validation (Collins et al., 2024; Fente et al., 2023; Kassahun et al., 2024). This context is relevant to Bima City because nutritional status and anaemia indicators can be obtained practically through antenatal care and have the potential to be used as initial components of risk screening. The relative novelty of this study lies not in identifying anaemia or nutritional status as individual risk factors, but in integrating both into a local probabilistic model, evaluating discrimination and calibration, and correcting for optimism using bootstrap internal validation.

METHOD

Study Design, Population, and Unit of Analysis

This study used a quantitative analytical approach within a framework for developing a preterm birth risk prediction model. The target population, as specified in the proposal, consisted of pregnant women in Bima City. The main study variables were maternal nutritional status and anaemia status, while the outcome was preterm birth. Planned covariates included maternal age, parity, and antenatal care (ANC) visits. At the model-development analysis stage, A total of 44 pregnant women were selected using purposive sampling. Participants with complete data on birth status, KEK status, anaemia, maternal age, parity, and antenatal care (ANC) visits were included in the analysis. The outcome variable was binary-coded as 1 for preterm birth and 0 for non-preterm birth. Preterm birth was defined as birth at a gestational age of less than 37 weeks, whereas birth at a gestational age of 37 weeks or more was categorised as non-preterm. The data collection period, sampling technique, and initial sample size before complete-case analysis need to be stated in the final manuscript according to the study implementation documents because this information was not explicitly reported in the proposal and Chapter IV on which this article is based.

Study Variables and Operational Definitions

The dependent variable was preterm birth status. The main independent variables were maternal nutritional status and anaemia. In the proposal, nutritional status was planned to be measured using body mass index (BMI) and/or mid-upper arm circumference (MUAC), while anaemia status was based on haemoglobin (Hb) levels. In the analytical dataset used to develop the core model, nutritional status was represented by chronic energy deficiency (KEK) status and coded as 1 for KEK and 0 for non-KEK; anaemia status was coded as 1 for anaemia and 0 for normal Hb levels. Covariates comprised maternal age in years, parity, and number of ANC visits. Parity was grouped as 0, 1-3, and more than 3, while ANC visits were grouped as fewer than 6 and 6 or more according to the analytical dataset structure. The BMI/MUAC and Hb cut-off values used to define KEK and anaemia categories were not explicitly stated in the available source documents; therefore, these values must be completed in the final manuscript based on the study instruments and data collection protocol rather than assumed.

Study Stages

The study stages followed the workflow formulated in the proposal, from population identification to the generation of risk classification outputs. This workflow was retained in the article so that the model development process could be traced systematically, as shown in Figure 1.

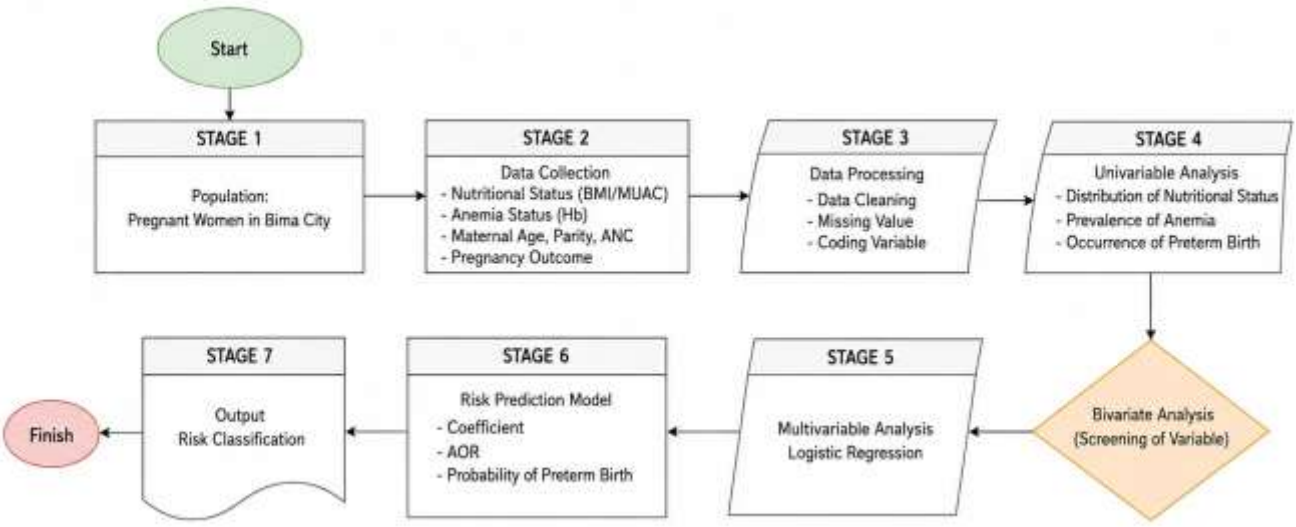


Figure 1. Stages of the research method

Stage 1 - Identification of the population and variables. The study population was defined as pregnant women in Bima City. At this stage, the preterm birth outcome, the main predictors of nutritional/KEK status and anaemia, and the covariates of maternal age, parity, and ANC visits were specified.

Stage 2 - Data collection. The collected data included nutritional status (BMI/MUAC), anaemia status based on Hb testing, maternal age, parity, history of ANC visits, and pregnancy outcome. These data formed the raw study dataset.

Stage 3 - Data processing. The dataset underwent completeness checks, data cleaning, missing-value handling, consistency verification, and variable coding. The model-development analysis then used complete cases with complete data for the main variables and covariates.

Stage 4 - Univariate and bivariate analyses. Univariate analysis was used to describe the distribution of nutritional status, prevalence of anaemia, maternal characteristics, and occurrence of preterm birth. Bivariate analysis served to screen the initial association of each predictor with the outcome and, in accordance with the proposal protocol, variables with $p<0,25$ were considered candidates for multivariable analysis.

Stage 5 - Multivariable analysis. Logistic regression was used to assess the simultaneous contributions of predictors, control for confounding factors, obtain regression coefficients, and calculate adjusted odds ratios (AORs).

Stage 6 - Risk prediction model development. Regression coefficients were used to construct the logit equation and estimate the individual probability of preterm birth. The main study model combined KEK status and anaemia in accordance with the proposal focus.

Stage 7 - Model evaluation and output. The model was evaluated in terms of discrimination, calibration, classification-threshold performance, and internal validation. The study outputs were risk probabilities and prediction classifications, which at this stage were positioned as pilot/exploratory results rather than final clinical cut-offs.

Data Collection and Processing

Data collection was conducted according to the predefined study variables, including maternal nutritional status, anaemia status, maternal age, parity, antenatal care (ANC) visits, and pregnancy

outcome. Nutritional status was assessed using BMI/MUAC indicators, while anaemia status was determined based on haemoglobin (Hb) levels. Birth outcome was determined based on gestational age at delivery. Before analysis, the data underwent completeness checks, *cleaning*, consistency verification, and variable coding to minimise errors in the analytical process. The final analysis used a complete-case dataset, including only respondents with complete data for all main variables and covariates included in the modelling. This approach was used to maintain consistency of the analytical units across all stages of model development and evaluation.

Univariate and Bivariate Analyses

Univariate analysis was used to describe respondent characteristics and the distribution of each variable. Continuous variables were presented as means and standard deviations or medians as appropriate, whereas categorical variables were presented as frequencies and percentages. The distributions of the outcome, KEK status, anaemia, parity, and ANC visits were analysed to describe the initial pattern in the dataset. Bivariate analysis was used to assess the initial association between each factor and preterm birth. Fisher's Exact Test was used for categorical variables because the number of preterm birth events and several cell frequencies were relatively small. Crude odds ratios and 95% confidence intervals were calculated as measures of association before adjustment. Maternal age and pre-pregnancy BMI were compared by birth status using the Mann-Whitney U test. In accordance with the proposal protocol, p<0,25 was used as the screening criterion for multivariable candidates; meanwhile, KEK and anaemia remained evaluated in the core model because both were main predictors specified a priori based on the study objectives. Multivariable analysis used binary logistic regression because the study outcome was dichotomous. The general model formulation planned in the proposal included nutritional status, anaemia, maternal age, parity, and ANC visits as predictors of the probability of preterm birth:

$$Logit(P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (1)$$

where X_1=nutritional status/KEK, X_2=anaemia status, X_3=maternal age, X_4=parity, and X_5=ANC visits. The adjusted odds ratio was calculated as AOR = exp(β). A p-value <0,05 was used as the threshold for statistical significance in interpreting coefficients, while also considering the 95% confidence interval and estimate stability. During the analysis, modelling was conducted in stages. Model 1 included KEK status only, Model 2 included anaemia only, Model 3 combined KEK and anaemia, while Model 4 included KEK, anaemia, maternal age, parity categories, and ANC. Model 3 was designated as the core model because it directly represented the two main study predictors and was more parsimonious relative to the number of available events. Model 4 was treated as a diagnostic analysis because there were only five preterm birth events for approximately six parameters, resulting in a high risk of overfitting and quasi-complete separation. The core model equation was expressed as $Logit(P) = \beta_0 + \beta_1(KEK) + \beta_2(Anemia)$. Individual probabilities were calculated using the following logistic function:

$$P = 1 / [1 + exp(-Logit(P))] \quad (2)$$

This probability represents the estimated likelihood of preterm birth for a given combination of KEK and anaemia status. At this stage of the study, the probability was treated as an exploratory predicted probability and had not yet been used as a basis for clinical decision-making.

Performance Evaluation and Classification Threshold

Model evaluation included discrimination and calibration. Discrimination was assessed using the receiver operating characteristic curve and area under the curve (ROC-AUC). Calibration was assessed using the Brier Score and calibration plot; the Hosmer-Lemeshow test was used as an additional analysis. Interpretation of calibration did not rely solely on Hosmer-Lemeshow because the number of preterm birth events was very limited. To evaluate the use of probabilities for classification, several thresholds were compared based on sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, F1-score, and Youden Index. The threshold with the highest Youden Index was used as the exploratory threshold. Because the

ultimate purpose is related to pregnancy screening, sensitivity and false negatives remained important considerations when assessing threshold suitability.

Internal Validation and Model Stability

Internal validation was performed using bootstrap resampling to estimate performance optimism in the development dataset. A total of 1.000 bootstrap replications were planned and 996 replications were successfully analysed. In each replication, the model was re-estimated to obtain the distribution of coefficients and changes in performance. Mean AUC optimism was then used to calculate the optimism-corrected ROC-AUC, and a similar evaluation was performed for the Brier Score. Resampling approaches such as bootstrap are relevant for estimating optimism and model-performance stability in development data (Coley et al., 2023; Collins et al., 2024). Bootstrap was selected because a simple train-test split is less efficient for a dataset with only five preterm birth events. The width of the bootstrap coefficient distribution was used as an additional indicator of parameter stability.

RESULT

Respondent Characteristics

A total of 44 respondents met the complete-case analysis criteria. Five respondents experienced preterm birth, resulting in a prematurity proportion of 11,36% in the analytical dataset. The mean maternal age was 31,80 +/- 6,27 years, and the mean pre-pregnancy BMI was 22,11 +/- 2,76 kg/m2. Seven respondents (15,9%) were categorised as having KEK and 5 respondents (11,4%) had anaemia.

Table 1.
Respondent characteristics by birth status

Characteristics	Category	Total	Non-Preterm	Preterm
Maternal age (years)	Mean +/- SD	31,80 +/- 6,27	31,92 +/- 6,51	30,80 +/- 4,27
Pre-pregnancy BMI	Mean +/- SD	22,11 +/- 2,76	22,26 +/- 2,68	21,01 +/- 3,47
KEK status	No	37 (84,1%)	33 (84,6%)	4 (80,0%)
KEK status	Yes	7 (15,9%)	6 (15,4%)	1 (20,0%)
Anaemia status	Normal	39 (88,6%)	37 (94,9%)	2 (40,0%)
Anaemia status	Anaemia	5 (11,4%)	2 (5,1%)	3 (60,0%)
Parity	0	12 (27,3%)	10 (25,6%)	2 (40,0%)
Parity	1-3	30 (68,2%)	27 (69,2%)	3 (60,0%)
Parity	>3	2 (4,5%)	2 (5,1%)	0 (0,0%)
ANC visits	<6 visits	34 (77,3%)	32 (82,1%)	2 (40,0%)
ANC visits	>=6 visits	10 (22,7%)	7 (17,9%)	3 (60,0%)

The most contrasting distribution was observed for anaemia status. Of the five respondents with anaemia, three (60,0%) experienced preterm birth. In contrast, only two of the 39 respondents with normal haemoglobin status experienced preterm birth. In the KEK group, one of seven respondents (14,3%) experienced preterm birth, whereas in the non-KEK group there were four cases among 37 respondents (10,8%).

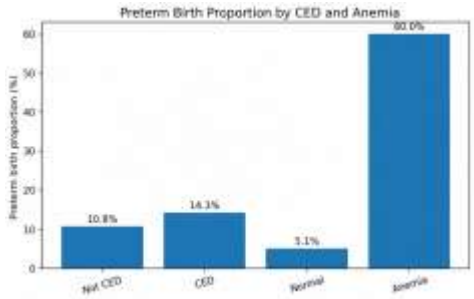


Figure 2. Proportion of preterm births by KEK and anaemia status

Bivariate Analysis

Bivariate analysis was performed to assess the association of each factor with preterm birth. Categorical variables were analysed using Fisher's Exact Test, whereas differences in maternal age and pre-pregnancy BMI between the preterm and non-preterm groups were analysed using the Mann-Whitney U Test.

Table 2.
Association of maternal factors with preterm birth

Variable	Crude OR	95% CI	p
KEK status	1,375	0,130-14,529	1,000
Anaemia	27,750	2,821-272,945	0,007
ANC >=6 visits	6,857	0,959-49,037	0,069

Based on table 2, anaemia showed a statistically significant association with preterm birth in the bivariate analysis. Women with anaemia had substantially higher odds of preterm birth than those with normal haemoglobin status. However, the very wide confidence interval indicates considerable uncertainty in the estimate. KEK did not show a significant association. Likewise, maternal age and pre-pregnancy BMI did not differ significantly between the preterm and non-preterm groups, with p-values of 0,617 and 0,318, respectively. ANC visits >=6 times showed a crude OR of 6,857 for prematurity with p=0,069. This finding should not be interpreted as indicating that a higher number of ANC visits causes prematurity. One possible explanation is that women with high-risk pregnancies received more intensive care contacts, so the association may reflect reverse causation or confounding by indication.

Univariat analysis

Univariable logistic regression analysis was conducted to evaluate the individual association between each maternal factor and preterm birth. Each predictor was entered separately into the model to estimate its crude effect before adjustment for other variables. The analysis included chronic energy deficiency (CED), anemia, maternal age, antenatal care (ANC), parity, and pre-pregnancy BMI.

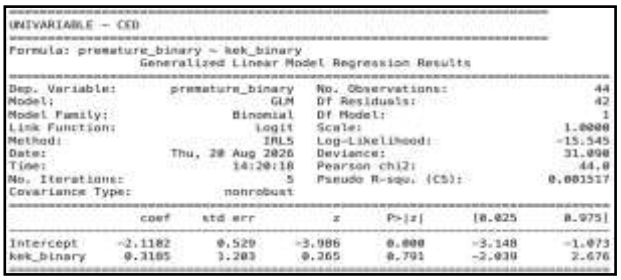


Figure (a) Univariate-KEK

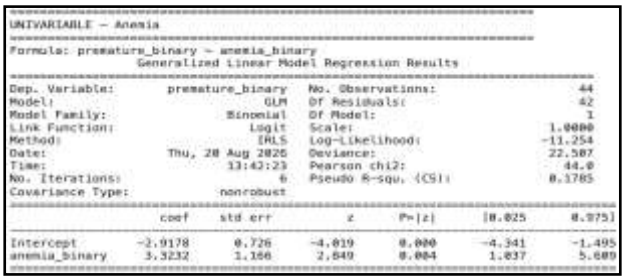


Figure (b) Univariate-Anemia

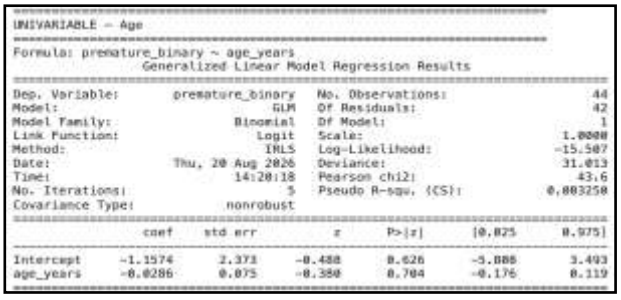


Figure (c) Univariate-Age

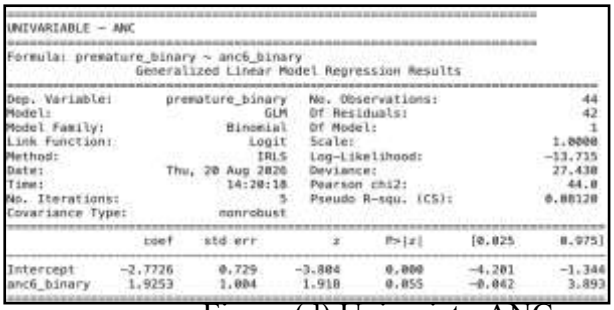


Figure (d) Univariate-ANC

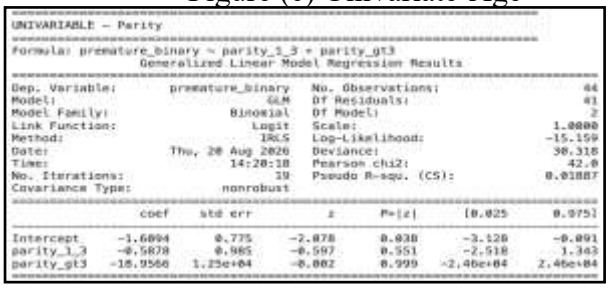


Figure (e) Univariate-Parity

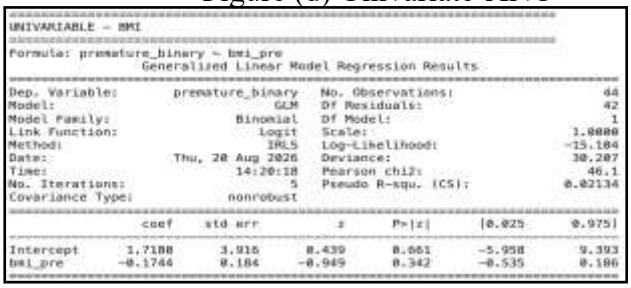


Figure (a) Univariate-BMI

Figure 3. Univariable Logistic Regression Results for Maternal Factors Associated with Preterm Birth: (a) Chronic Energy Deficiency (CED), (b) Anemia, (c) Maternal Age, (d) Antenatal Care (ANC), (e) Parity, and (f) Pre-pregnancy Body Mass Index (BMI).

Figure 3, illustrates the results of the univariable logistic regression analysis examining the association between each maternal factor and preterm birth. Anemia showed the strongest and statistically significant association with preterm birth (B = 3.3232, p = 0.004), while ANC showed a borderline association (B = 1.9253, p = 0.055). In contrast, chronic energy deficiency (CED), maternal age, parity, and pre-pregnancy BMI were not significantly associated with preterm birth (p > 0.05). These findings were used to support the selection of candidate variables for subsequent multivariable logistic regression analysis.

Logistic Regression Model Development

In Model 1, KEK yielded an OR of 1,375 (95% CI 0,130-14,528; p=0,791). In Model 2, anaemia yielded an OR of 27,750 (95% CI 2,821-272,934; p=0,004). After KEK and anaemia were entered simultaneously in Model 3, the KEK coefficient was 0,2041 and the anaemia coefficient was 3,3171.

Table 3.
Core preterm birth prediction model

Parameter	Beta	AOR	95% CI	p
Intercept	-2,952	-	-	<0,001
KEK	0,204	1,226	0,069-21,715	0,889
Anaemia	3,317	27,579	2,800-271,677	0,004

Logit(P) = -2,9517 + 0,2041(KEK) + 3,3171(Anemia)

Table 3 shows that anaemia produced a much larger change in log-odds than KEK. To improve interpretability, the theoretical probabilities based on combinations of the two predictors were calculated from the logistic equation.

Table 4.
Predicted probabilities by combination of KEK and anaemia status

KEK	Anaemia	Predicted probability
No	No	4,97%
Yes	No	6,02%
No	Yes	59,03%
Yes	Yes	63,86%

Table 4 presents probability values showing that, in this pilot model, changes in probability were driven primarily by anaemia status. KEK without anaemia increased the probability only from approximately 4,97% to 6,02%, whereas anaemia increased the probability to approximately 59,03%. Model 4, which included KEK, anaemia, age, parity, and ANC, produced several extreme coefficients and large standard errors and therefore was not used as the main source of AOR estimates.

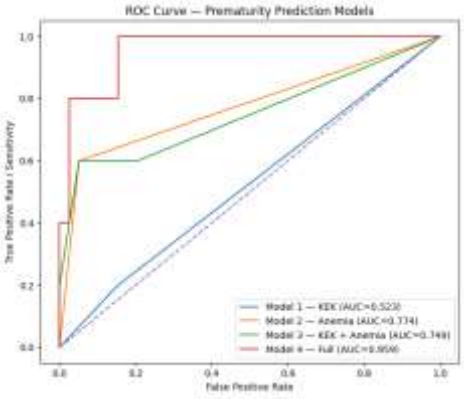
Model Discrimination and Calibration

Model performance was evaluated to assess the extent to which the developed prediction model could distinguish respondents who experienced preterm birth from those who did not and generate predicted probabilities consistent with the observed outcomes. Discrimination was evaluated using ROC-AUC, while calibration was assessed using the Brier Score, *calibration plot*, and the Hosmer-Lemeshow test.

Table 5.
Performance of the logistic regression models

Model	ROC-AUC apparent	Brier Score
Model 1 – KEK	0,523	0,101
Model 2 - Anaemia	0,774	0,070
Model 3 - KEK + anaemia	0,749	0,070
Model 4 - Full adjusted	0,959	0,049

Based on table 5, the apparent ROC-AUC of Model 3 was 0,7487, indicating initial discriminatory ability in the development dataset. However, apparent performance does not yet represent the performance that can be expected in new data. The Brier Score for Model 3 was 0,0697. The Hosmer-Lemeshow test yielded p=1,000, but this value cannot be used as the sole basis for concluding that calibration is good because there were only five preterm birth events.



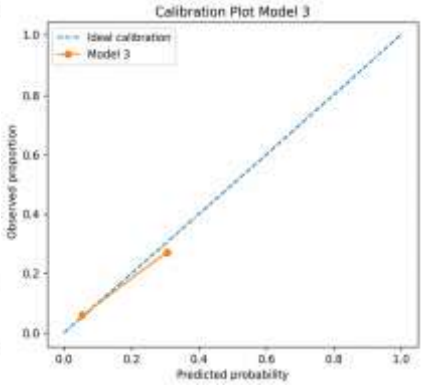


Figure (a). ROC curves for Models 1 to 4 *Figure (b). Calibration plot for Model 3*
Figure 4. Model Evaluation

Classification Threshold and Performance

The threshold based on the Youden Index was 0,5903. At this threshold, the confusion matrix yielded 3 true positives, 2 false positives, 37 true negatives, and 2 false negatives. Sensitivity was 60,0%, specificity 94,9%, PPV 60,0%, NPV 94,9%, accuracy 90,9%, and F1-score 0,600.

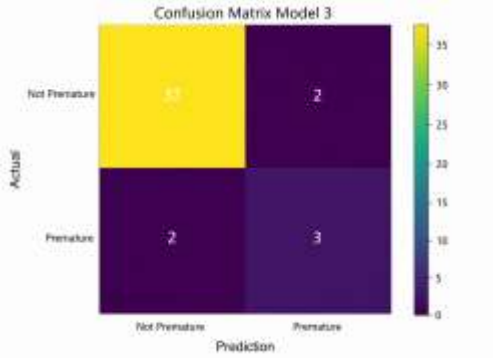


Figure 5. Confusion matrix for Model 3 at the Youden threshold

Figure 5 shows that the model correctly classified most non-preterm respondents, but it remained limited in detecting preterm cases. Of the 5 preterm cases, only 3 were correctly identified, while the other 2 were not detected. Thus, the model sensitivity of 60,0% indicates that its ability to identify preterm cases remains limited. Although overall accuracy reached 90,9%, this value should be interpreted cautiously because the outcome distribution was imbalanced, with 39 non-preterm respondents and only 5 preterm respondents. Therefore, in a screening context, sensitivity deserves greater attention than overall accuracy.

Model Validation

Validation was performed to assess model stability and estimate possible *optimism* in the performance of the developed model. The validation process used a bootstrap approach because this method allows model evaluation without splitting the dataset into training and test data, making it more suitable for the limited sample size and number of preterm birth events.

Table 6.
Validation results for Model 3

Indicator Validation	**Value**
ROC-AUC apparent	0,7487
Mean optimism AUC	0,0763
Optimism-corrected ROC-AUC	0,6725
Brier Score apparent	0,0697
Optimism-corrected Brier Score	0,089
Successful bootstrap	996/1.000

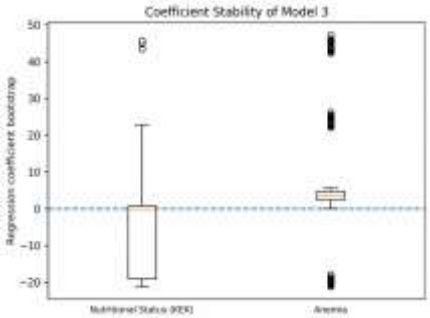


Figure 6. Stability of KEK and anaemia coefficients based on bootstrap resampling in Model 3

Based on Table 6 and Figure 6, the internal validation results for Model 3 showed that after correction for optimism, the AUC decreased from 0,7487 to 0,6725. This decline indicates that part of the discriminatory ability observed in the development dataset was attributable to optimism. The bootstrap distributions of the coefficients were also very wide and consistent with the small number of events and respondents in the anaemia and KEK categories. Therefore, the model should be regarded as a pilot/exploratory model.

DISCUSSION

This study produced three main findings. First, anaemia showed the strongest association with preterm birth, both in the bivariate analysis and after being combined with KEK status. Second, KEK did not show a clear statistical association in this sample. Third, the two-predictor model demonstrated moderately promising initial discrimination in the development dataset, but its performance declined after optimism correction and the model parameters showed substantial instability.

Anaemia as the Dominant Predictor

In the core model, anaemia yielded an AOR of 27,579 with a 95% CI of 2,800-271,677. Statistically, this finding indicates that women with anaemia had higher odds of preterm birth after accounting for KEK status. The direction of this association is consistent with international and regional literature reporting increased prematurity among pregnancies complicated by anaemia (Beressa et al., 2024; Khan et al., 2024; Khezri et al., 2023; Wang et al., 2025). However, the AOR of 27,579 in this study is much larger than most effect estimates reported in studies with large samples. This difference should not be interpreted as evidence that the biological effect of anaemia in Bima City is dozens of times greater. Only five women had anaemia, and three of them experienced preterm birth. The combination of very small frequencies and a highly concentrated outcome distribution can produce extreme odds ratios and very wide confidence intervals. Therefore, the substantive message of this study is not that anaemia increases risk by exactly 27,6 times, but rather that anaemia shows a strong and consistent predictive signal that warrants re-evaluation in a larger sample. This finding is relevant to public health because anaemia can be detected through ANC services and some of its causes can be prevented or managed through nutritional and maternal-care approaches (González-Fernández et al., 2024; Wang et al., 2025).

KEK Status and Prematurity

KEK status yielded an AOR of 1,226 with p=0,889 and a confidence interval of 0,069-21,715. Inferentially, this study was not able to demonstrate a clear independent association between KEK and prematurity. This finding should not be simplified into the conclusion that nutritional status is irrelevant. Only seven respondents were categorised as having KEK and only one of them experienced preterm birth, resulting in very limited statistical power to detect small to moderate effects. Larger studies indicate that the relationship between maternal nutritional status and prematurity is complex and may vary according to anthropometric indicators, parity, maternal disease, and pregnancy characteristics (González-Fernández et al., 2024; Jiang et al., 2025; Sawadogo et al., 2024). In addition, nutritional status in the main model was represented as the categorical KEK variable. Categorisation results in the loss of biological variation that may be

present in continuous measurements. In future studies, the use of MUAC and/or BMI as continuous variables should be considered provided that measurement quality is adequate.

Interpretation of the Combined KEK and Anaemia Model

The combination of KEK and anaemia was selected because both indicators can reflect aspects of maternal nutritional vulnerability and are available in antenatal care. A simple model is also more likely to be implemented than a model requiring many variables that are difficult to collect. The probability results show that the current pilot model is driven almost entirely by anaemia. The baseline probability for women without KEK and without anaemia was approximately 4,97%. KEK without anaemia increased the probability only to 6,02%, whereas anaemia without KEK increased it to approximately 59,03%. This pattern explains why the ROC-AUC of Model 2, which used anaemia alone (0,774), was slightly higher than that of Model 3 (0,749). The addition of KEK to this dataset did not provide clear incremental predictive value. Nevertheless, a decision to permanently exclude KEK cannot be made on the basis of 44 respondents. A larger sample may produce different estimates and allow the contribution of KEK to be evaluated with greater precision.

Discrimination, Calibration, and Optimism

The apparent ROC-AUC of 0,7487 indicates that Model 3 could distinguish, to some extent, between women with and without preterm birth in the development dataset. However, after bootstrap correction, the ROC-AUC decreased to 0,6725. The decrease of approximately 0,076 is methodologically important because evaluation on the same dataset used for model development tends to produce overly optimistic performance (Coley et al., 2023; Collins et al., 2024). This finding also explains why Model 4, with an apparent AUC of 0,959, should not be regarded as the best model. The very high apparent performance occurred in the context of only five events and many parameters. Extreme coefficients, large standard errors, and the absence of preterm cases in the parity >3 category are indicators of instability and possible separation. Thus, the more complex model actually has a greater risk of overfitting.

Screening Threshold

The Youden threshold of 0,5903 yielded a sensitivity of 60,0% and specificity of 94,9%. Mathematically, this threshold is very close to the probability generated for women with anaemia but without KEK, approximately 59,03%. Consequently, the classification of Model 3 in this dataset was strongly influenced by the presence of anaemia. From a maternal-health screening perspective, a sensitivity of 60% is not adequate as a basis for implementation. Two of the five preterm births were not identified. The consequences of false negatives in maternal screening may be more important than an increase in false positives. Therefore, the threshold of 0,5903 should be regarded as an exploratory data-based threshold rather than a clinical cut-off.

Implications for Public Health Services

The study findings have relevant implications for strengthening ANC services. Anaemia and nutritional status can be used not to replace clinical assessment, but as initial components of risk flagging. In practical terms, a parsimonious model based on routinely available variables has the potential to be developed more readily into a screening tool than a model that depends on costly examinations. If the model is subsequently validated using a larger dataset, risk probabilities may help identify women who require further assessment, more intensive nutritional counselling, anaemia evaluation, reinforcement of treatment adherence, and pregnancy monitoring. However, this study does not yet support the use of the model as a clinical decision-making tool. At present, it is best positioned as a proof-of-concept based on local data.

Strengths and Limitations of the Study

A strength of this study is that it goes beyond association analysis. The study developed a probability equation, evaluated discrimination, measured the Brier Score, performed threshold analysis, and used bootstrap resampling to estimate optimism. This approach is consistent with

modern practice in the development and reporting of prediction models (Coley et al., 2023; Collins et al., 2024). The main limitation was that there were only five preterm birth events. This number is very small for obtaining stable parameter estimates. A second limitation was that only five respondents had anaemia and seven had KEK, resulting in very wide confidence intervals and unstable bootstrap coefficient distributions. In the context of prediction modelling, a limited number of events also increases vulnerability to overfitting and unstable estimates (Coley et al., 2023; Collins et al., 2024).

A third limitation was that anaemia status in the analytical dataset was available as a categorical variable, so variation in haemoglobin levels could not be fully utilised. Similarly, the use of KEK as a categorical variable may result in loss of information if continuous MUAC measurements are available. Fourth, Model 4 was at risk of quasi-complete separation because the parity >3 category contained no preterm events. Fifth, the study has not yet undergone external validation. The model cannot be assumed to perform similarly across different facilities, periods, or populations without further evaluation.

CONCLUSION

This study successfully developed a pilot model for predicting preterm birth based on KEK status and anaemia among pregnant women in Bima City, with the equation Logit(P) = -2,9517 + 0,2041(KEK) + 3,3171(Anemia). Anaemia was the predictor showing the strongest signal, with an AOR of 27,579 (95% CI 2,800-271,677; p=0,004), whereas KEK did not show a significant independent association, with an AOR of 1,226 (95% CI 0,069-21,715; p=0,889). The model yielded an apparent ROC-AUC of 0,7487, which decreased to 0,6725 after optimism correction using bootstrap. At the exploratory threshold of 0,5903, sensitivity reached 60,0% and specificity 94,9%. These findings indicate that the model was successfully developed technically, but its performance is not yet sufficiently stable for use as a clinical screening tool.

Anaemia can be prioritised as an important indicator in further model development, while the contribution of KEK should be re-evaluated using a larger sample and more informative measures of nutritional status. Further development should increase the number of preterm birth events, consider haemoglobin and anthropometric indicators as continuous variables, update coefficient estimates, evaluate calibration more robustly, determine thresholds based on clinical objectives, and perform external validation. The model developed in this study should therefore be positioned as a pilot or exploratory model based on the local population.

REFERENCES

Adigama, I. P., Sayang, N., Ngurah, G., Yuliastina, N., Pasek, I. M., & Gauthama, S. (2025). Maternal Oxygen Transport Capacity and Nutritional Reserves: Anemia and Mid-Upper Arm Circumference (MUAC) as Independent Predictors of Low Birth Weight in the Indonesian Highlands. In Community Medicine and Education Journal. https://doi.org/10.37275/cmej.v7i1.832

Beressa, G., Whiting, S. J., Kuma, M. N., Lencha, B., & Belachew, T. (2024). Association between anemia in pregnancy with low birth weight and preterm birth in Ethiopia: A systematic review and meta-analysis. In PLoS ONE. https://doi.org/10.1371/journal.pone.0310329

Chun, R. P. C., Chan, H. G., Lim, G. Y. S., Kanagalingam, D., Partana, P., Tan, K. H., Teoh, T. G., & Tan, I. (2025). Preterm birth trends and risk factors in a multi-ethnic Asian population: A retrospective study from 2017 to 2023, can we screen and predict this? In Annals of the Academy of Medicine, Singapore. https://doi.org/10.47102/annals-acadmedsg.202518

Coley, R. Y., Liao, Q., Simon, N., & Shortreed, S. (2023). Empirical evaluation of internal validation methods for prediction in large-scale clinical data with rare-event outcomes: A case study in suicide risk prediction. In BMC Medical Research Methodology. https://doi.org/10.1186/s12874-023-01844-5

Collins, G. S., Moons, K., Dhiman, P., Riley, R., Beam, A., Calster, B. V., Ghassemi, M., Liu, X., Reitsma, J. B., Smeden, M. van, Boulesteix, A., Camaradou, J., Celi, L., Denaxas, S.,

Denniston, A., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., … Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. In British medical journal. https://doi.org/10.1136/bmj-2023-078378

Fente, B. M., Asaye, M. M., Tesema, G., & Gudayu, T. (2023). Development and validation of a prognosis risk score model for preterm birth among pregnant women who had antenatal care visit, Northwest, Ethiopia, retrospective follow-up study. In BMC Pregnancy and Childbirth. https://doi.org/10.1186/s12884-023-06018-1

González-Fernández, D., Muralidharan, O., Neves, P. A., & Bhutta, Z. (2024). Associations of Maternal Nutritional Status and Supplementation with Fetal, Newborn, and Infant Outcomes in Low-Income and Middle-Income Settings: An Overview of Reviews. In Nutrients. https://doi.org/10.3390/nu16213725

Huang, C., Long, X., Ven, M. van der, Kaptein, M., Oei, S. G., & Heuvel, E. R. van den. (2024). Predicting preterm birth using electronic medical records from multiple prenatal visits. In BMC Pregnancy and Childbirth. https://doi.org/10.1186/s12884-024-07049-y

Jiang, Y., Wang, X., Wu, L., Huang, X., & Cao, X. (2025). Effects of Maternal Prepregnancy Nutritional Status on Pregnancy Outcomes. In Emergency Medicine International. https://doi.org/10.1155/emmi/1502902

Kassahun, E. A., Gebreyesus, S., Tesfamariam, K., Endris, B., Roro, M., Getnet, Y., Hassen, H. Y., Brusselaers, N., & Coenen, S. (2024). Development and validation of a simplified risk prediction model for preterm birth: A prospective cohort study in rural Ethiopia. In Scientific Reports. https://doi.org/10.1038/s41598-024-55627-z

Khan, Z. A., Khail, S. K., Qureshi, A., & Ahmad, P. (2024). The association between maternal anemia and preterm birth: A case-control study. In Journal of Shifa Tameer-e-Millat University. https://doi.org/10.32593/jstmu/vol7.iss1.318

Khezri, R., Salarilak, S., & Jahanian, S. (2023). The association between maternal anemia during pregnancy and preterm birth. In Clinical Nutrition ESPEN. https://doi.org/10.1016/j.clnesp.2023.05.003

Liu, Y., Liu, J., & Shen, H. (2024). Machine learning model‐based preterm birth prediction and clinical nomogram: A big retrospective cohort study. In International journal of gynaecology and obstetrics: The official organ of the International Federation of Gynaecology and Obstetrics. https://doi.org/10.1002/ijgo.16036

Nadhiroh, S., Hasugian, A. R., Nurhayati, Muthiah, A., Nadhira, A., & Putri, P. A. (2024). MODEL DEVELOPMENT FOR ANEMIA PREDICTION IN PREGNANCY. In Clinical Epidemiology and Global Health. https://doi.org/10.1016/j.cegh.2024.101654

Sawadogo, W., Tsegaye, M., Gizaw, A., Newland, H. L., & Adera, T. (2024). Maternal Prepregnancy Body Mass Index and Risk of Preterm Birth: The Role of Weight Gain during Pregnancy, Race, and Ethnicity. In American Journal of Perinatology. https://doi.org/10.1055/a-2494-2080

Wang, R., Xu, S., Hao, X., Jin, X., Pan, D., Xia, H., Liao, W., Yang, L., & Wang, S. (2025). Anemia during pregnancy and adverse pregnancy outcomes: A systematic review and meta-analysis of cohort studies. In Frontiers in Global Women's Health. https://doi.org/10.3389/fgwh.2025.1502585

Yu, Q.-Y., Lin, Y., Zhou, Y.-R., Yang, X.-J., & Hemelaar, J. (2024). Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms. In Frontiers Big Data. https://doi.org/10.3389/fdata.2024.1291196

Yu, X., Nie, H., Yang, W., Tang, J., Huang, X., & Zhao, Y. (2026). Association of hemoglobin trajectories during pregnancy with preterm birth and non-reassuring fetal status: A retrospective cohort study using latent growth mixture modeling. In BMC Pregnancy and Childbirth. https://doi.org/10.1186/s12884-026-09688-9