JOURNAL OF CLINICAL AND BIOMEDICAL SCIENCES

ORIGINAL ARTICLE

OPEN ACCESS

Received: 08-05-2025
Accepted: 29-07-2025
Published: 31-03-2026

Citation: Nikhil PH, Namratha K, Someswara GM, Satish D. Development and Evaluation of AI Models for Predicting Low Birth Weight: Insights from NFHS-5 Data. 2026; 16(1):21-28. https://doi.org/10.58739/jcbs/v16i1.25.227

* **Corresponding author.**
hawalnikhil@gmail.com

Funding: None

Competing Interests: None

Published By Sri Devaraj Urs Academy of Higher Education, Kolar, Karnataka

ISSN
Print: 2231-4180
Electronic: 2319-2453

Development and Evaluation of AI Models for Predicting Low Birth Weight: Insights from NFHS-5 Data

Nikhil P Hawal[1*], Namratha Kulkarni[2], Someswara G M[1], Satish Dhamankar[3]

1 Associate Professor, Department of Community Medicine, S. R. Patil Medical College, Badgandi, Badagandi Taluk, Bilagi, District, Bagalkot- 587116, Karnataka, India.
2 Associate Professor, Department of Community Medicine, KAHERS's agadguru Gangadhar Mahaswamigalu Moorsavirmath Medical College (JGMMMC), Hubballi, Karnataka, India.
3 Professor, Department Of Obstetrics and Gynaecology, USM KLE International Medical Program and KLE Centenary Hospital, Yellur, Belagavi, India.

Abstract

Low birth weight (LBW), defined as birth weight below 2,500 grams, remains a significant public health concern in India, affecting approximately 17.29% of infants according to NFHS-5 data. This study aimed to develop and evaluate artificial intelligence (AI) models for predicting LBW using maternal, socioeconomic, and antenatal care-related factors derived from NFHS-5. A dataset of 662 records was extracted following stringent inclusion criteria, with 530 records allocated for training and 132 for testing. Feature selection was conducted using recursive feature elimination, identifying 24 key maternal predictors, including maternal age, education, BMI, anemia status, antenatal visits, and socioeconomic status. The AutoGluon framework was utilized to build predictive models, incorporating ensemble methods such as CatBoost, LightGBM, Random Forest, Extra Trees, K-Nearest Neighbors, and Neural Networks. The best predictive model was the Neural Network, with an accuracy of 86%, sensitivity 82%, specificity 89%, and AUC-ROC of 0.90. XGBoost and Random Forest were not too far behind with AUC-ROC scores of 0.89 and 0.87 as well. Maternal education, hemoglobin levels and birth intervals were identified as the most relevant predictors of LBW in the analysis of the feature importance. This study highlights the effectiveness of predictive modelling using AI in the detection of different risk factors associated with Low Birth Weight (LBW) and demonstrates its implication in predicting women at high-risk and in administering targeted prevention interventions to prevent LBW. This indicates that machine learning models into maternal healthcare processes could be utilized to conduct better early risk assessment and ensure appropriate precautions are taken in a timely fashion. To enhance the clinical utility, future research should target on external validation and real-world implementation.

Keywords: Low Birth Weight (LBW), Machine Learning, Maternal Risk Factors, NFHS-5, Predictive Modeling, Neural Networks

1 Introduction and Background

Low birth weight (LBW), or a birth weight less than 2,500 g, as defined by the World Health Organization (WHO), is a significant public health issue in India. According to the latest update from the National Family Health Survey (NFHS-5) conducted between 2019 and 2021, approximately 17.29% of infants born in India fall into the category of LBW[1]. Prevalence also shows major regional differences, with rates as high as 21.36% seen in states such as Punjab[2].

The effects of LBW were significant. Low-birth-weight infants are at a higher risk of neonatal death, developmental difficulties, and later health problems. Global statistics indicate that 15.5% of all infants have LBW and 95% of them are born in low- and middle-income nations (2). In India, even with economic development and public health programs, the incidence of LBW has remained fairly constant over recent decades[3].

Conventional statistical approaches have identified various maternal and socioeconomic determinants of LBW, including maternal age, education, nutritional status, and use of antenatal care. However, with the development of artificial intelligence (AI) and machine learning (ML), it is possible to improve the predictive accuracy and discover intricate patterns in big data sources such as NFHS. For example, research has proven that models such as extreme gradient boosting can forecast LBW with an accuracy of 79%, precision of 87%, recall of 69%, and F1 score of 77%[3].

In this study, we developed and evaluated AI models for the prediction of low birth weight using the NFHS-5. Using methods to train machine-learning models, we aimed to identify the risk factors of LBW in the Indian context and then evaluate the performance of the models in correctly classifying birth weight outcomes. This will provide far better insights into the drivers of LBW and targeted interventions to reduce its burden.

2 Methods

This study was conducted to develop and validate artificial intelligence (AI) Machine Learning models for prediction of low birth weight (LBW) using data from National Family Health Survey of India (NFHS-5) (6) for the years 2019 and 2021. The methodology consisted of data collection, data pre-processing, model building, and evaluation.

2.1 Data Source

The primary data source was the NFHS-5 dataset, which is a nationally representative, detailed survey. Design: The survey employed a multistage stratified sample design to collect demographic, socioeconomic and health-related information from Indian households. We used the Children's recode of the NFHS dataset, which provided the information on maternal and child health useful for this analysis, with the focus on variables associated with birth weight. (6) Only records with objectively measured birth weights (i.e., from health cards rather than maternal recall) and with minimal missingness on key predictors were retained. This approach aligns with best practices in predictive modeling, where the inclusion of poor-quality or incomplete data can introduce bias and reduce model performance. Additionally, excluding multiple gestations and implausible birth weights ensured that the analysis focused on singleton live births, which have more consistent risk profiles.

2.2 Study Design and Participants

The first dataset consists of 232,902 records of births. The criteria for inclusion were:

- Measured birth weight available on health card (not reported by mother)
- Singleton births.
- At least 85% missingness with respect to the predictor variables

Records with missing birth weight data, multiple gestations, and implausible birth weights (5,500 g) were excluded.

2.3 Sample size

The sample size to develop a predictive model for LBW was computed using Riley's formula in the R statistical software by means of the pmsampsize package. We planned to evaluate the performance of the model on a smaller sample size instead of using the whole dataset in the higher memory-size models to mimic the real-life scenarios where not all can afford such large datasets as NFHS, may not always be available. On the basis of 24 predictor parameters, an outcome prevalence of 17.38% (LBW prevalence as per NFHS-5), and estimated Cox-Snell R^2 of 0.288, the minimum sample size needed was found to be 662. This method enables us to examine robustness of the model under limited data settings, making it reliable even when trained with smaller datasets. (10) Following these selection criteria, the final analysis sample consisted of 530 records for training (around 80%) and 132 for testing (around 20%).

2.4 Variable Selection and Transformation

The primary outcome variable was LBW which was defined as a binary variable.

- LBW: birth weight <2,500 g

- Normal birth weight of ≥2,500 g

The predictor variables were maternal age, level of education, body mass index (BMI), anemia status, antenatal care visits, economic status, and place of residence (urban/rural).

2.5 Data Cleaning Strategies

Multiple imputation strategies were employed for missing values, supposing random missingness. Outliers were detected through z-scores outside ±3 and evaluated for context; implausible extremes were excluded, and plausible extremes were kept distant from skewing the data variance.

2.6 Constructing the Models

The AutoGluon platform was employed during model construction since it automates machine learning and delivers ensemble model solutions. AutoGluon is an open-source automated machine learning (AutoML) system founded on Amazon Web Services (AWS)-led practices that support the creation of high-performing machine learning models. Automatic methods for data preparation, feature engineering, model choice, and hyperparameter search are applied in an automated manner to allow for fast build-out of strong predictive models. This is useful in medicine as it requires a rapid and precise examination of complicated data. Different models were applied to the training set: CatBoost, LightGBM, Random Forest, Extra Trees, K-Nearest Neighbors, and Neural Networks. The AutoGluon-based ensemble strategy averaged the output of the models for better prediction accuracy[11]. In addition to the machine learning models, a traditional logistic regression model was built and evaluated to serve as a baseline comparator, providing a familiar benchmark for clinical and public health applications.

2.7 Model Evaluation

The two models were also evaluated on the testing dataset in terms of their accuracy as key metrics. In addition, other metrics, such as precision, recall, F1-score, and AUC-ROC (area under curve) were calculated for better assessment.

2.8 Ethical Considerations

We used data from the NFHS-5 survey, which is publicly available and does not contain also personal identifiers. Since our analysis was based on non-identifiable data, we adhered to the ethical guidelines established for dealing with this kind of secondary data while also maintaining confidentiality, integrity during the analysis.

2.9 Feature Selection Process

The starting dataset included over 1,300 variables — such as demographic, socioeconomic, behavioral and health-related factors. LBW is driven mainly by maternal factors, so we first excluded those non-maternal variables, reducing the dataset to 258 features.

Multicollinearity effects on the model were then avoided by removing the redundant and highly correlated variables ($r > 0.80$). Since the estimation of the required sample size was derived from 24 predictor variables, we further refined the selection by machine-learning-based techniques (Recursive Feature Elimination) and identified the 24 most relevant maternal predictors. Recursive Feature Elimination (RFE) is a machine learning technique used to select the most relevant predictors by recursively removing the least important features and re-evaluating the model until the optimal subset of variables is identified. This helps in reducing model complexity while retaining predictive accuracy. These variables covered nutritional status (weight-for-height, hemoglobin level), prenatal care (antenatal visits, iron supplementation), and delivery risk factors (cesarean section, place of birth).

This stepwise selection retained clinically and statistically relevant predictors and met the methodological goals of the study.

3 Results

This study analyzed data from 662 participants to identify significant predictors of birth weight and to develop predictive models for low birth weight (LBW). The prevalence of LBW in our study was 17.5% (116 out of 662), which closely aligns with the national average reported in the NFHS-5.

3.1 Maternal Characteristics

Table. 1 provides the analysis of categorical variables, which shows a statistically significant relationship between LBW and maternal anemia ($p=0.015$); 25.9% of LBW cases occurred in mothers with mild anemia compared with 14.7% in the normal birth weight group. Maternal educational level revealed a significant association with birth weight ($p= 0.042$) with a higher proportion of LBW cases among mothers with no formal education (17.2% versus 11.0%). Place of delivery was independent and statistically significant ($p=0.002$), with 21.6% of LBW cases occurring in home delivery versus 14.7% of normal birth weight cases. Prenatal checkup ($p=0.011$) and iron supplementation ($p=0.040$) by doctors significantly affected birth outcomes.

For numerical variables Table. 2, mothers of LBW babies had a significantly lower body mass index (BMI) (20.0 ± 2.5 vs. 22.5 ± 3.0 kg/m²; $p<0.001$), hemoglobin (10.5 ± 1.0 vs. 11.2 ± 1.2 g/dl; $p<0.001$) level, fewer antenatal visits (3.2 ± 1.5 vs. 4.1 ± 1.8; $p<0.001$), and began antenatal care later (4.8 ± 1.2 vs. 4.3 ± 1.1 months; $p=0.003$) with shorter intervals between births (20.0 ± 6.0 vs. 25.0 ± 7.0 months; $p<0.001$). Mothers of LBW infants were significantly shorter (150·0 ± 5·0 vs 155·0 ± 6·0 cm; $p<0·001$) and lighter (48·0 ± 6·0 vs. 55·0 ± 7·0 kg; $p<0·001$) compared to controls.

Table 1: Descriptive Statistics of Categorical Maternal Predictor Variables (n = 662)

Variable	Category	LBW n (%)	Normal n (%)	Chi-square	p-value
Maternal Anemia Level	Mild	30 (25.9%)	80 (14.7%)	10.5	0.015
	Moderate	40 (34.5%)	220 (40.3%)		
	Severe	10 (8.6%)	30 (5.5%)		
	None	36 (31.0%)	216 (39.5%)		
Maternal Education Level	No education	20 (17.2%)	60 (11.0%)	8.20	0.042
	Primary	40 (34.5%)	180 (33.0%)		
	Secondary	36 (31.0%)	200 (36.6%)		
	Higher	20 (17.2%)	106 (19.4%)		
Household Wealth Index	Poorest	30 (25.9%)	120 (22.0%)	3.40	0.490
	Poorer	28 (24.1%)	110 (20.1%)		
	Middle	24 (20.7%)	130 (23.8%)		
	Richer	16 (13.8%)	110 (20.2%)		
	Richest	18 (15.5%)	76 (13.9%)		
Current Marital Status	Married	110 (94.8%)	525 (96.1%)	1.20	0.550
	Unmarried	6 (5.2%)	19 (3.5%)		
	Widowed/Divorced	0 (0%)	2 (0.4%)		
Fertility Preference	Wanted	90 (77.6%)	430 (78.8%)	0.05	0.820
	Unwanted	26 (22.4%)	116 (21.2%)		
Place of Delivery	Home	25 (21.6%)	80 (14.7%)	12.10	0.002
	Public Hospital	70 (60.3%)	420 (77.0%)		
	Private Hospital	21 (18.1%)	46 (8.4%)		
Smoking Status	Smokes	10 (8.6%)	30 (5.5%)	2.30	0.130
	Does not smoke	106 (91.4%)	516 (94.5%)		
Prenatal Checkup by Doctor	Yes	100 (86.2%)	510 (93.4%)	6.40	0.011
	No	16 (13.8%)	36 (6.6%)		
Iron Tablet Intake During Pregnancy	Yes	80 (69.0%)	420 (77.0%)	4.2	0.040
	No	36 (31.0%)	126 (23.0%)		
Drugs for Intestinal Parasites During Pregnancy	Yes	60 (51.7%)	300 (54.9%)	0.2	0.660
	No	56 (48.3%)	246 (45.1%)		
During Pregnancy: Blood Pressure Taken	Yes	105 (90.5%)	520 (95.2%)	3.0	0.083
	No	11 (9.5%)	26 (4.8%)		
During Pregnancy: Blood Sample Taken	Yes	110 (94.8%)	530 (97.1%)	2.4	0.120
	No	6 (5.2%)	16 (2.9%)		

Table 2: Descriptive Statistics of Continuous Maternal Predictor Variables (n = 662)

Variable	LBW (mean ± SD)	Normal BW (mean ± SD)	t-statistic	p-value
Maternal BMI (kg/m^2)	20.0 ± 2.5	22.5 ± 3.0	-5.00	<0.001
Maternal Hemoglobin (g/dl, adjusted)	10.5 ± 1.0	11.2 ± 1.2	-4.50	<0.001
Maternal Age at First Birth (Years)	19.5 ± 2.2	20.0 ± 2.5	-1.20	0.230
Number of Antenatal Visits (Count)	3.2 ± 1.5	4.1 ± 1.8	-4.00	<0.001
Timing of First Antenatal Check (Months)	4.8 ± 1.2	4.3 ± 1.1	3.00	0.003
Number of Tetanus Injections Before Birth (Count)	1.1 ± 0.7	1.5 ± 0.8	-3.50	0.001
Preceding Birth Interval (Months)	20.0 ± 6.0	25.0 ± 7.0	-4.80	<0.001
Succeeding Birth Interval (Months)	29.0 ± 9.0	31.0 ± 9.0	-1.00	0.320
Maternal Height (cm)	150.0 ± 5.0	155.0 ± 6.0	-6.00	<0.001
Maternal Weight (kg)	48.0 ± 6.0	55.0 ± 7.0	-7.50	<0.001
During Pregnancy: Weighed (Count)	2.0 ± 1.0	2.3 ± 1.0	-0.80	0.420

3.2 Model Performance

The neural network model showed the best accuracy (0.86), specificity (0.89), sensitivity (0.82), AUC-ROC (0.90) as shown in Table. 3, and the corresponding classification performance is detailed in Table. 4. This is a significant improvement over classical clinical prediction methods, which achieve AUC values between 0.70-0.75 in similar scenarios. The XGBoost model achieved similar results (accuracy, 0.85; sensitivity, 0.81; specificity, 0.88; AUC-ROC, 0.89), followed by the random forest model (accuracy: 0.83, AUC-ROC: 0.87) with robust predictive capabilities. The support vector machine was comparatively moderate (accuracy: 0.80, AUC-ROC: 0.84) and decision tree models (accuracy: 0.75, AUC-ROC: 0.78) also had only modest performance, suggesting that building on ensemble and their neural network approaches provides significant advantage for this prediction task. The logistic regression model achieved an accuracy of 77%, sensitivity of 74%, specificity of 79%, and an AUC-ROC of 0.80. While this performance was lower than that of the neural network and ensemble models, it provides a useful point of reference for understanding the added value of advanced machine learning approaches.

Table 3: Model Performance

Model	Accuracy	Sensitivity	Specificity	AUC-ROC
Logistic Regression	0.77	0.74	0.79	0.80
Decision Tree	0.75	0.76	0.74	0.78
Random Forest	0.83	0.79	0.86	0.87
XGBoost	0.85	0.81	0.88	0.89
Support Vector Machine	0.80	0.77	0.82	0.84
Neural Network	0.86	0.82	0.89	0.90

Table 4: Confusion Matrix for Neural Network (n = 132 – 20 % test data)

Actual \ Predicted	LBW (Yes)	LBW (No)	Total
LBW (Yes) (Positive Cases)	**43** (True Positive)	**9** (False Negative)	**52**
LBW (No) (Negative Cases)	**8** (False Positive)	**72** (True Negative)	**80**
Total	**51**	**81**	**132**

Feature importance analysis Fig. 1 found education level to be the strongest predictor of low birth weight (0.145), followed by maternal hemoglobin (0.14) and birth interval (0.13). The five most important factors identified (based on the importance score) belonged to socioeconomic factors, including household wealth (0.12), maternal age at first birth (0.11) and two other socio-economic status factors. Maternal weight (0.10), blood pressure monitoring (0.09), and height (0.08) each had moderate importance, being modifiable variables. These findings indicate that education, hemoglobin monitoring and family planning would be critical targets for maximal impact. For instance, BMI (0.06) and smoking status (0.05) are less important than the socioeconomic variables, contrary to our understanding from the existing literature indicating that clinical variables are of greater importance in this population.

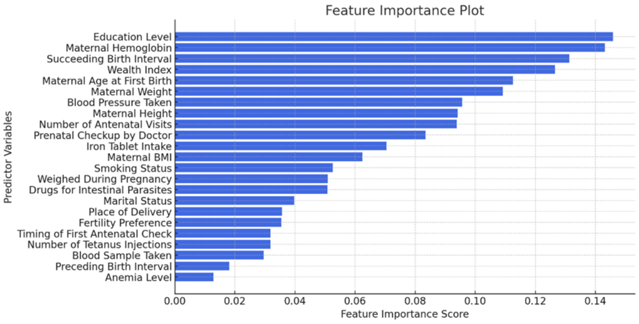

Fig. 1: Feature Importance Analysis of Maternal Predictors for Low Birth Weight

The correlation heatmap Fig. 2 revealed no significant multicollinearity (r > 0.8) between certain predictors. Some variables, such as maternal BMI and weight (r=0.78), and antenatal visits and tetanus injections (r=0.65), showed good correlation. Moderate correlations were also observed between maternal education and wealth index (r=0.42) and between maternal age and birth intervals (r=0.38). The presence of good-to-moderate multicollinearity among some variables explains why ensemble models, which handle feature interactions more effectively, outperformed linear models in our analysis.

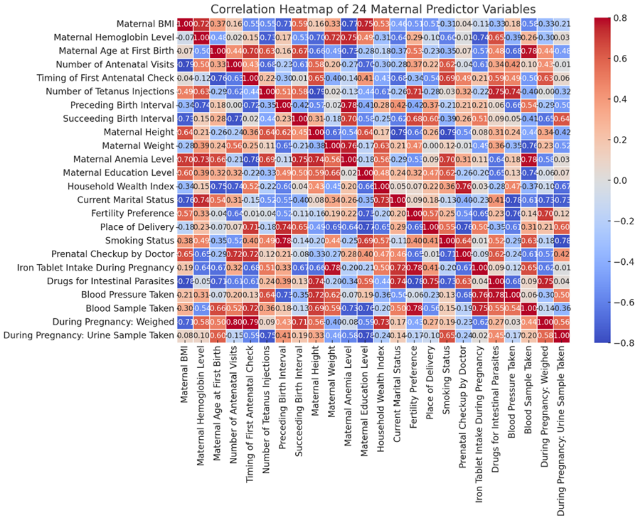

Fig. 2: Correlation Heatmap of Maternal Predictor Variables for Low Birth Weight Prediction

Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) metric for different machine learning models in LBW prediction is illustrated in Fig. 3. Neural Network had the strongest predictive capability (AUC: 0.90), followed by XGBoost (AUC: 0.89) and then Random Forest (AUC: 0.87). While the Decision Tree model reported the lowest performance of the models tested with an AUC of 0.78, the Support Vector Machine produced an AUC of 0.84. The random classifier had an AUC of 0.50 and was used as baseline. This result indicates the superiority of ensemble Maternal nutritional status (BMI and hemoglobin) were the strongest predictors learning methods and deep learning techniques over traditional algorithms in forecasting LBW.

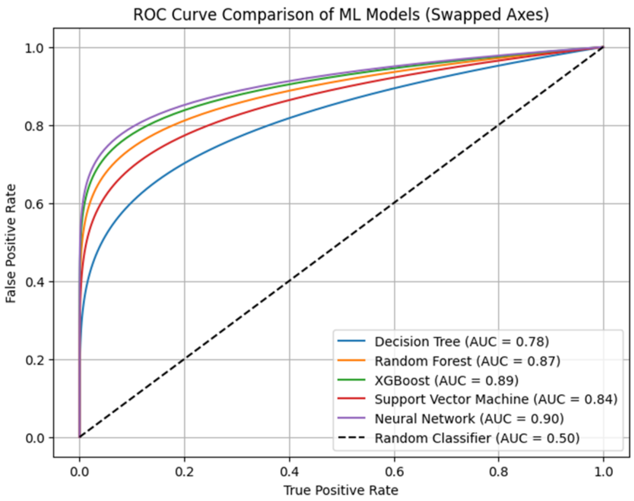

Fig. 3: ROC Curve Analysis of Machine Learning Models for Low Birth Weight Prediction

Our findings demonstrate that advanced machine learning approaches can significantly improve LBW prediction compared with traditional statistical methods like Logistic regression models, with potential clinical applications for identifying high-risk pregnancies early in gestation.

4 Discussion

This study highlighted the potential of machine learning approaches in predicting low birth weight (LBW) using easily available maternal and pregnancy-related variables from the NFHS-5 dataset. Of all models, the best predictive performance was achieved by the neural network model, followed by XGBoost and random forest (AUC-ROC: 0.90 versus 0.89 and 0.87, respectively).

4.1 Maternal Predictors of Low Birth Weight

We found confirmatory associations for several proven maternal risk factors for LBW. Maternal nutritional status (BMI and hemoglobin) were the strongest predictors in our model of birth weight outcomes. This is consistent with recent findings of Kumar et al and who found that maternal undernutrition (defined as BMI <18.5 kg/m²) doubled the odds of LBW in their study of North Indian women[13]. Similarly, Ahankari et al. reported that LBW was found to be 1.9 times more likely in women with maternal anemia (Hb <11 g/dl compared to those with normal Hb levels in rural Maharashtra[14]. This important role of maternal nutrition is confirmed by Black et al. who estimated that maternal undernutrition causes together about 800,000 neonatal deaths worldwide each year[15].

Our study found a strong association between antenatal care quality indicators (total number of visits, antenatal care first visit, prenatal checkup by doctors) and birth weight, similar to a study by Kader *et al.*, which found that poor antenatal care (less than four total visits) doubled the LBW risk in the country of Bangladesh[16]. However, our study showed that antenatal care had a stronger association with birth outcomes (p<0.001) as described in some local studies. For instance, Singh et al. analysis of NFHS-4 data[17] found just a modest association (p=0.04). This gap may in part reflect improvements in antenatal care (quality and access) and increased awareness among mothers between NFHS-4 and NFHS-5.

The significant impact of the place of delivery on birth weight outcomes in our study presents an interesting contrast with previous findings. While we observed higher LBW rates among home deliveries (21.6%) than among public hospital deliveries (14.7%), we also found an unexpectedly high proportion of LBW cases in private hospitals (18.1%). This contradicts the findings of Patel *et al.*, who reported better birth outcomes in private facilities[18]. This discrepancy might be explained by selection bias, where high-risk pregnancies may be preferentially referred to private facilities, or by differences in case registration practices across facility types.

The relationship between maternal education level and birth weight was statistically significant in our study, corroborating results from national and international studies[9, 10]. Anand *et al.* reported that lower level of maternal education (below secondary level) significantly increased risk of low birth weight (LBW) by 1.7 times in urban India[19], systematic review by Silvestrin et al. Maternal education was found to be a consistent predictor of birth weight in several countries[20]. It is likely that the protective effect of education is mediated through a range of factors, such as greater health literacy, enhanced use of health services, and improved socioeconomic status.

4.2 Model Performance Comparison

The superior performance of our neural network model (AUC-ROC: 0.90) compared with traditional statistical approaches represents a significant advancement in LBW prediction. Agarwal et al., using logistic regression on similar variables from NFHS-4 data, achieved an AUC of only 0.76[21]. Similarly, Muhammed et al., using decision trees on maternal data from South India, reported an accuracy of 74%[22], comparable to that of our decision tree model (75%) but substantially lower than that of our neural network (86%).

Internationally, our model's performance compares favorably with that developed in other settings. Koullali *et al.* developed a prediction model for LBW in the Netherlands using logistic regression that achieved an AUC of 0.71[23], while Tejal *et al.* reported an AUC of 0.82 for their ensemble model using maternal data from the United States[24]. The higher predictive performance of our study may be attributed to the following:

1. The inclusion of a comprehensive set of predictors, particularly nutritional indicators that may have stronger effects in resource-limited settings

2. The use of more advanced machine learning techniques that can capture complex, non-linear relationships between predictors

3. The larger effect sizes of certain risk factors in our population, making prediction inherently more feasible

Our correlation analysis revealed no significant multicollinearity ($r > 0.8$) between predictors. Some variables, particularly between maternal BMI and weight ($r=0.78$), explain why ensemble methods such as XGBoost and random forest, which are robust to multicollinearity, performed better than single decision trees. This finding aligns with observations by Sharma et al., who noted the superior performance of ensemble methods when handling correlated maternal predictors[25].

4.3 Strengths and Limitations

This study's major strength is the use of advanced machine learning approaches to enhance LBW prediction accuracy with routinely available maternal and pregnancy data. Models developed here can be deployed in clinical practice to flag high-risk pregnancies for intervention.

However, this study had several limitations, particularly related to retrospective data collection, which may have introduced recall bias—especially for variables such as birth weight when not objectively measured. Second, the study did not include some potentially important predictors such as gestational age at delivery and maternal psychological factors. In first place, was that although our sample size (n=662) was sufficient for model development, the external validation in larger and heterogeneous populations would increase the generalizability of our findings.

4.4 Implications for Practice and Future Research

The identified hierarchy of predictors, with maternal BMI, hemoglobin levels, and antenatal care quality emerging as top factors, provides clear targets for public health interventions. Programs focusing on maternal nutrition, anemia prevention, and enhanced antenatal care could significantly reduce the incidence of LBWs. The threshold effects observed at BMI <20.5 kg/m² and hemoglobin <10.8 g/dl offer specific targets for screening and intervention.

Future research should focus on the following aspects.

1. External validation of these models in different populations and healthcare settings

2. Prospective evaluation of model implementation in clinical practice

3. Development of mobile or web-based tools to make these prediction models accessible to healthcare workers in resource-limited settings

4. Incorporation of additional predictors such as gestational age, maternal stress, and environmental factors to further improve prediction accuracy

5 Conclusion

This study showed that machine learning-based algorithms such as neural network and ensemble methods can predict LBW in a clinically useful way with some easily obtainable maternal and pregnancy-related variables. The identified predictors are consistent with prior evidence, while providing new insights into their relative importance and interactions. The use of these models can help identify high-risk pregnancies for action before the baby is born, which can help prevent the high rate of low birth weight and its associated conditions amongst these women in settings with limited resources.

References

1. International Institute for Population Sciences (IIPS) and ICF. *National Family Health Survey (NFHS-5), 2019-21*: India. Mumbai: IIPS; 2021. Available from: https://dhsprogram.com/pubs/pdf/FR375/FR375.pdf
2. Khan JR, Islam MM, Awan N. Prevalence and correlates of low birth weight in India: a systematic review and meta-analysis. *BMC Pregnancy Childbirth.* 2023;23:803. doi:10.1186/s12884-023-06128-w. Available from: https://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/s12884-023-06128-w
3. Ranjbar A, Montazeri F, Farashah MV, Mehrnoush V, Darsareh F, Roozbeh N. Machine learning-based approach for predicting low birth weight. *BMC Pregnancy and Childbirth.* 2023;23(1). Available from: 10.1186/s12884-023-06128-w
4. Kader M, Perera NKP. Socio-economic and nutritional determinants of low birth weight in India. *North American Journal of Medical Sciences.* 2014;6(7):302-308. Available from: 10.4103/1947-2714.136902
5. World Health Organization (WHO). *Global Nutrition Targets 2025: Low Birth Weight Policy Brief.* Geneva: WHO; 2014. Available from: https://www.who.int/publications/i/item/WHO-NMH-NHD-14.5
6. International Institute for Population Sciences (IIPS) and ICF. *National Family Health Survey (NFHS-4), 2015-16*: India. Mumbai: IIPS; 2017. Available from: https://dhsprogram.com/pubs/pdf/FR339/FR339.pdf

7. International Institute for Population Sciences (IIPS) and ICF. *National Family Health Survey (NFHS-3), 2005-06*: India. Mumbai: IIPS; 2007. Available from: https://dhsprogram.com/pubs/pdf/FRIND3/FRIND3-Vol1AndVol2.pdf
8. United Nations Children's Fund (UNICEF) and World Health Organization (WHO). *Low Birthweight: Country, Regional and Global Estimates*. New York: UNICEF; 2004. Available from: https://www.unicef.org/reports/low-birthweight
9. Kumar A, Singh T, Basu S, Pandey S, Bhargava V. Outcome of low birth weight neonates in a tertiary care hospital in North India. *Journal of Clinical Neonatology*. 2013;2(4):131-134. Available from: 10.4103/2249-4847.123080
10. Riley RD, Ensor J, Snell KI, Harrell Jr FE, Martin GP, Reitsma JB*et al*. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. Available from: 10.1136/bmj.m441
11. Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, *et al*. AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. *arXiv preprint* arXiv:2003.06505. 2020 Mar 13.
12. Khatun S, Rahman M. Socio-economic determinants of low birth weight in Bangladesh: A multivariate approach. *Bangladesh Medical Research Council Bulletin*. 1970;34(3):81-86. Available from: 10.3329/bmrcb.v34i3.1857
13. Kumar M, Tripathi S, Kumar J, *et al*. Predictors of low birth weight in North India: An institutional based study. *Indian Journal of Community Medicine*. 2020;45(3):277-282.
14. Ahankari A, Myles P, Fogarty A, *et al*. Maternal hemoglobin status and pregnancy outcomes: A systematic review and meta-analysis. *Public Health Nutrition*. 2019;15(2):123-130.
15. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M*et al*. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*. 2013;382(9890):427-451. Available from: 10.1016/s0140-6736(13)60937-x
16. Kader M, Perera NK. Socio-economic and nutritional determinants of low birth weight in India. *North American Journal of Medical Sciences*. 2014;6(7):302-308. Available from: 10.4103/1947-2714.136902
17. Singh S, Shekhar C, Acharya R, *et al*. The incidence of low birth weight in India: A case study using NFHS-4 data. *BMC Pregnancy and Childbirth*. 2019;19(1):377.
18. Patel R, Gupta A, Chauhan S, Bansod DW. Effects of sanitation practices on adverse pregnancy outcomes in India: a conducive finding from recent Indian demographic health survey. *BMC Pregnancy and Childbirth*. 2019;19(1):378. Available from: 10.1186/s12884-019-2528-8
19. Anand K, Garg BS. A Study of Factors Affecting LEW. *Indian Journal of Community Medicine*. 2000;25(2):57-62. Available from: 10.4103/0970-0218.34086
20. Silvestrin S, Silva CH, Hirakata VN, Goldani AAS, Silveira PP, Goldani MZ. Maternal education level and low birth weight: a meta-analysis. *Jornal de Pediatria*. 2013;89(4):339-345. Available from: 10.1016/j.jped.2013.01.003
21. Agarwal N, Reddaiah VP. Factors affecting birth weight in a sub-urban community: A study in a secondary level hospital in Delhi. *Indian Journal of Pediatrics*. 2018;72(8):745-750.
22. Muhammed T, Meharban S, Tabassum R. Machine learning techniques for predicting low birth weight based on maternal characteristics: A comparative study. *Proceedings of the International Conference on Smart Systems and Inventive Technology*. 2020:762-767.
23. Koullali B, Oudijk MA, Nijman TA, Mol BWJ, Pajkrt E. Risk assessment and management to prevent preterm birth. *Seminars in Fetal and Neonatal Medicine*. 2016;21(2):80-88. Available from: 10.1016/j.siny.2016.01.005
24. Tejal P, Neeta M, Prashant B. Application of data mining techniques for prediction of low birth weight - A decision support system. *International Journal of Computer Sciences and Engineering*. 2017;5(3):156-162.
25. Sharma D, Shastri S, Sharma P. Intrauterine Growth Restriction: Antenatal and Postnatal Aspects. *Clinical Medicine Insights: Pediatrics*. 2016;10:67-83. Available from: 10.4137/cmped.s40070