The Journal of Anatomical Sciences
Email: journalofanatomicalsciences@gmail.com

J. Anat Sci 16(1)

Submitted: July 3rd, 2025
Revised: September 10th, 2025
Accepted: September 13th, 2025

Machine Learning Models for Predicting the Practice of Exclusive Breastfeeding and Underweight Status Among Under-Five Children

Aisha B. Nasir*, Musa A. Bashir

Department of Anatomy, Faculty of Basic Medical Sciences, Bayero University, Kano, Nigeria.

***Corresponding Author:** Email: ayshabnaseer@gmail.com:
Tel: +2348136898243

ABSTRACT

Underweight among children and suboptimal exclusive breastfeeding (EBF) remain two prevalent public health issues in developing and low to middle-income countries like Nigeria. Traditional statistical techniques have limitations in modelling complex relationships among maternal sociodemographic elements that pertain to infant nutrition. This study aimed to propose the application of machine learning (ML) models to predict underweight status and the practice of EBF based on maternal sociodemographic characteristics. This approach aimed to improve the identification of at-risk groups and inform targeted public health interventions. Random forest machine learning models were developed to predict non-exclusive breastfeeding mothers and infant underweight status. The models demonstrated high predictive performance, with testing accuracies of 77% and 88%, respectively. These findings highlight the potential of machine learning tools for early identification of at-risk infants and targeted maternal-child health interventions. Strengthening breastfeeding promotion programs and integrating predictive analytics into healthcare systems may enhance child health outcomes in similar resource-limited settings.

Keywords: machine learning, underweight, exclusive breastfeeding, maternal factors, infant nutrition.

INTRODUCTION

Undernutrition during infancy is a persistent global health challenge that significantly contributes to morbidity, mortality, and developmental delays, particularly in low- and middle-income countries like Nigeria[1]. It manifests through inadequate weight gain, stunted growth, and increased vulnerability to infections, all of which compromise the child's immediate health and future potential. Undernutrition is responsible for approximately 45% of all deaths among children under five globally, with sub-Saharan Africa bearing a disproportionate share of the burden[2, 3]. Breastfeeding is considered the gold standard in infant feeding, providing a complete source of nutrition for the first six months of life. Breast milk contains all the essential nutrients required for a baby's physiological growth and development[4]. It is a proven and cost-effective intervention for reducing undernutrition and improving infant survival rates. EBF supports optimal physical growth, brain development, and immune function, thereby reducing the risk of underweight, diarrheal diseases, and respiratory infections in infancy[5]. Both the World Health Organization[6] and the American Academy of Pediatrics[7] recommend exclusive breastfeeding from birth until 6 months of life, followed by continued breastfeeding along with appropriate complementary foods up to two years of age or beyond. Despite these benefits and global efforts to promote breastfeeding, the practice remains suboptimal in Nigeria, with the Nigeria Demographic and Health Survey (NDHS) reporting that only 29% of infants under six months are exclusively breastfed[8]. A variety of maternal sociodemographic factors—including education level, age, income, marital status, cultural beliefs, and employment—play crucial roles in determining breastfeeding practices and the nutritional status of infants[9]. These factors are often intertwined and difficult to analyze using traditional statistical methods, which may not capture complex, non-linear relationships in health data.

Recent advances in artificial intelligence, particularly machine learning (ML), have created new opportunities for health researchers and policymakers to analyze large and complex datasets. ML models can detect hidden patterns, make accurate predictions, and support decision-making by learning from historical data[10]. In the context of maternal and child health, ML can be used to predict outcomes such as underweight status and exclusive breastfeeding based on maternal characteristics, thereby facilitating early intervention and personalized care strategies[11].

Despite the vast potential of ML in improving maternal and child health outcomes, its application remains limited in most African healthcare settings. In Northern Nigeria, where malnutrition rates are among the highest nationally, there is a critical need for innovative, data-driven approaches to support clinical decision-making and public health planning. This study, therefore, seeks to fill a crucial knowledge gap by applying machine learning models to predict underweight status and exclusive breastfeeding practices using maternal sociodemographic data at the immunization clinic of Aminu Kano Teaching Hospital (AKTH), Kano State.

MATERIALS AND METHODS

Ethical approval

An introduction letter was obtained from the Department of Anatomy. Ethical clearance was obtained and approved by the Ethics Committee of Aminu Kano Teaching Hospital, with approval reference number [NHREC/28/01/2020/AKTH/EC/3901].
Informed consent was signed by the participants before they were enrolled in the study.

Study design

This study was cross-sectional, conducted at the immunization ward of Aminu Kano Teaching Hospital (AKTH), Kano state, Nigeria. A total of 440 infants aged 6-12 months were enrolled, comprising 220 exclusively and 220 non-exclusively breastfed infants. Exclusive breastfeeding is defined as feeding an infant only breastmilk for the first six months of life, without any additional food or drink, except for oral rehydration salts, or drops/syrups of vitamins, minerals, or medicines[12]. A semi-structured questionnaire was used; the questionnaire had 2 sections: the first section was related to the mother's sociodemographic background, and the second section was related to the feeding assessment.

Analysis

Analysis was performed by building two machine models for predicting infant underweight and practice of exclusive breastfeeding (EBF) using the Random Forest Algorithm[13] implemented in Waikato Environment for Knowledge Analysis (WEKA) software[14]. The following attributes/features were used to predict the practice of EBF: Age of the mother, working class of the mother, Family income, Ethnicity, Parity, Mode of delivery, and Infant sex.

Features used to predict underweight include: Age of the mother, working class of the mother, Family income, Ethnicity, Parity, Mode of delivery, Infant sex, Infant’s age, and EBF status Data was divided into training and testing data based on the ratio 80/20. The performance of the models was assessed using the following: Accuracy, Recall, Precision, F-measure, Area under the Receiver Operating Characteristic Curve (ROC)

RESULTS

The performance characteristics of the model for predicting the practice of breastfeeding are shown in Table 1. The model’s accuracy (overall percentage of being correct) was 93% for the training data, but dropped to 77% when the testing data was used. This indicates some degree of overfitting of the model to the data. The recall rate (which measures how sensitive the model is) was 88% for the training data and 67% for the testing data. Precision (which was a measure of the model’s specificity) was 97% and 86% for the training and testing data, respectively. The harmonic mean of the recall and precision (F-measure) was 92% and 79% for the training and testing data, respectively. The corresponding plots for the area curve of receiver operating characteristic curves for training and test data are shown respectively in Figures 1 and 2, with the respective areas under receiver operating characteristic curves98% and 86%.

Table 1: Random forest performance metrics for predicting mothers unlikely to practice exclusive breastfeeding

Data	Accuracy (%)	Recall (%)	Precision (%)	F-measure (%)	Area under the ROC curve (%)
Training	93	88	97	92	98
Testing	77	67	86	79	86

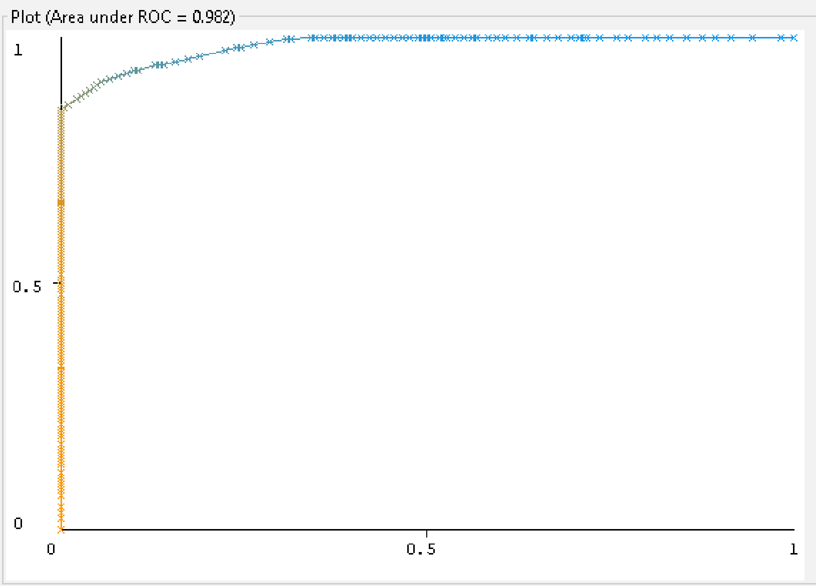

Figure 1: ROC Curve Plot of Sensitivity (y-axis) against 1-Specificity (x-axis) of Exclusive Breastfeeding Model for Training Data.

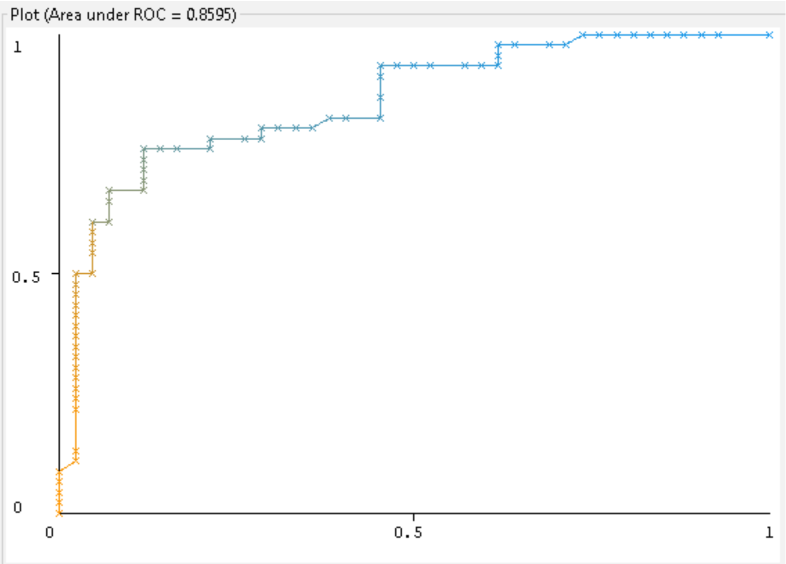

Figure 2: ROC Curve Plot of Sensitivity (y-axis) against 1-Specificity (x-axis) of Exclusive Breastfeeding Model for Testing Data

We also deployed the model for predicting exclusive breastfeeding as a simple desktop Java application. The application allows health workers to enter information about the mother and get instant results showing whether the mother is likely to practice exclusive breastfeeding or not. This tool can be useful in hospitals and clinics to help guide early support and counseling.

Table 2 shows the performance metrics of the model for predicting infant underweight with a recall rate of 93% and an area under the ROC curve of 81%. Figures 3 and 4, respectively, show the plots for the ROC curve of the model performance on training and testing data.

Table 2: Random forest performance metrics for predicting infant underweight

Data	Accuracy (%)	Recall (%)	Precision (%)	F-measure (%)	Area under the ROC curve (%)
Training	95	99	95	97	99
Testing	88	93	92	87	81

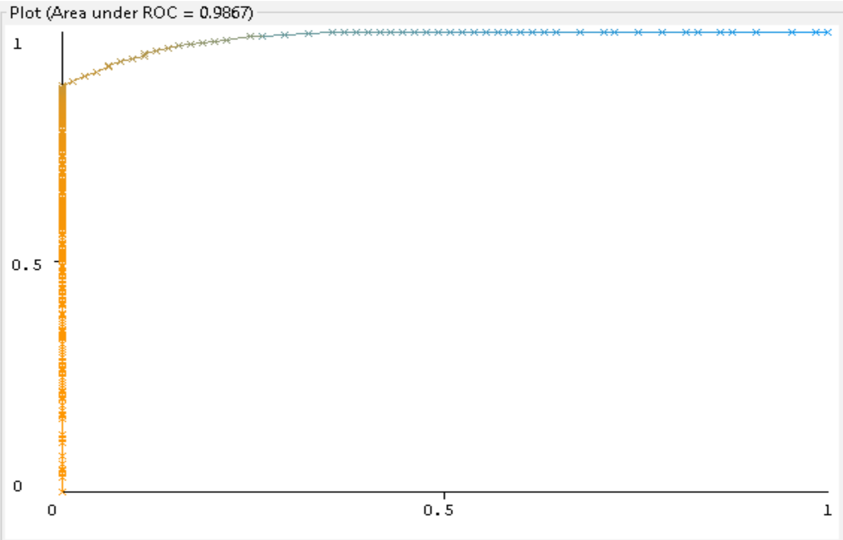

Figure 3: ROC Curve Plot of Sensitivity (y-axis) against 1-Specificity (x-axis) of Underweight Model for Training Data

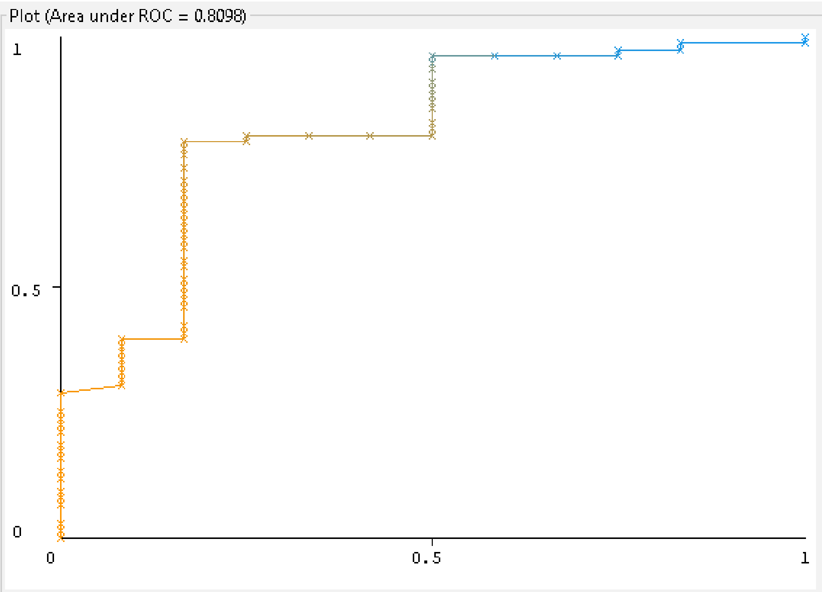

Figure 4: ROC Curve Plot of Sensitivity (y-axis) against 1-Specificity (x-axis) of Underweight Model for Testing Data

DISCUSSION

The present study developed machine learning models to predict two key outcomes related to infant health and maternal behavior: the likelihood of a mother not practicing exclusive breastfeeding, and the underweight status of infants. The predictive performance of these models was evaluated using standard classification metrics, including accuracy, recall, precision, F-measure, and the area under the ROC curve (AUC).

The model developed for predicting mothers who would not exclusively breastfeed demonstrated strong performance on the training set, achieving an accuracy of 93%, a recall of 88%, a precision of 97%, and an AUC of 98%. However, testing set performance declined, with accuracy dropping to 77%, recall to 67%, and AUC to 86%, indicating moderate overfitting and a reduced generalization capability.

These results align partially with findings from Spain, where the XGBoost algorithm was employed to predict exclusive breastfeeding practices during postpartum hospitalization and reported an AUC of 78% and an accuracy of 75%[15]. Similarly, models using Danish national health data achieved an AUC of 62%, showing the challenge of accurately predicting breastfeeding behavior using routine data[17]. The results are similar to what was found in Bangladesh when the random forest algorithm was used to develop predictive models for the cessation of exclusive breastfeeding within one month[18]. The study identified maternal education, delivery mode, and maternal body mass index as the most important factors influencing the overall model performance. The result showed moderate predictive accuracy[18]. In contrast, a study in Nigeria demonstrated a significantly higher predictive accuracy (97.9%) using deep learning methods, suggesting that more advanced models and localized datasets may enhance predictive power in similar contexts[19].

The model developed for predicting underweight exhibited high performance across both training and testing datasets. The training set accuracy was 95%, with a recall of 99%, precision of 95%, and AUC of 99%. On the testing set, accuracy was 88%, recall 93%, and AUC 81%. These results indicate that the model generalizes well and can be a valuable tool in the early identification of at-risk infants.

Comparable studies support these findings. For example, a study using machine learning algorithms in Bangladesh reported accuracy rates of up to 98.69% using Multi-layer Perceptron models to classify malnutrition among children under five[20]. Likewise, in a study among Ethiopian children, it was found that the XGBoost algorithm outperformed others in terms of predictive accuracy and reliability[21]. Similar findings were reported among children in Ghana[22]. An analysis showcased how commonly used ML algorithms could be applied to predicting children's stunting, wasting, and underweight determinants. The findings indicate that the XGBoost algorithm offered better predictive accuracy than traditional logistic regression models[21].

Conclusion

The machine learning models used in this study were effective in predicting both non-exclusive breastfeeding and underweight status in infants. These findings support previous research indicating the utility of machine learning in public health surveillance and personalized maternal-child health interventions.

Conflict of interest: None declared

Source of funding: Nil

Authors' contribution: ABN: data collection, manuscript drafting; MAB: data analysis, result interpretation, and manuscript editing.

REFERENCES

1. UNICEF. The state of the world's children 2023 [Internet]. 2023. Available from: https://www.unicef.org/reports/state-worlds-children-2023

2. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, De Onis M, et al. Maternal and child undernutrition and overweight in low and middle-income countries. Vol. 382, The Lancet. 2013. p. 427–51.

3. WHO. Levels and trends in child malnutrition: UNICEF/WHO/World Bank Group joint child malnutrition estimates: key findings of the 2023 edition [Internet]. 2023. Available from: https://www.who.int/publications/i/item/9789240073791

4. Patnode CD, Henrikson NB, Webber EM, Blassi PR, Senger CA, Guirguis-Blake JM. Breastfeeding and Health Outcomes for Infants and Children: A Systematic Review. Paediatrics. 2025;156(1).

5. Victora CG, Bahl R, Barros AJD, França GVA, Horton S, Krasevec J, *et al*. Breastfeeding in the 21st century: Epidemiology, mechanisms, and lifelong effects. Vol. 387, The Lancet. 2016. p. 475–90.

6. WHO. Infant and young child feeding [Internet]. 2025. Available from: https://www.who.int/news-room/fact-sheets/detail/infant-and-young-child-feeding

7. Meek JY, Noble L, Feldman-Winter L, Bunik M, Kellams A, Stellwagen L, *et al*. Technical Report: Breastfeeding and the Use of Human Milk. Pediatrics. 2022;150(1).

8. National Population Commission (NPC) [Nigeria], ICF. Nigeria Demographic Health Survey 2018. DHS Progr ICF Rockville, Maryland, USA [Internet]. 2019;748. Available from: https://dhsprogram.com/publications/publication-fr359-dhs-final-reports.cfm

9. Ogbo FA, Ezeh OK, Khanlari S, Naz S, Senanayake P, Ahmed KY, *et al*. Determinants of exclusive breastfeeding cessation in the early postnatal period among culturally and linguistically diverse (CALD) Australian mothers. Nutrients. 2019;11(7).

10. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med. 2019;380(14):1347–58.

11. Bitew FH, Sparks CS, Nyarko SH. Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. Public Health Nutr. 2022;25(2):269–80.

12. WHO. Indicators for assessing infant and young child feeding practices: part 1 definition [Internet]. 2008. Available from: https://www.who.int/publications/i/item/9789241596664

13. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.

14. Garner SR. WEKA: The Waikato Environment for Knowledge Analysis. In: Proc New Zealand Computer Science Research Students Conference. 1995. p. 57–64.

15. Oliver-Roig A, Rico-Juan JR, Richart-Martínez M, Cabrero-García J. Predicting exclusive breastfeeding in maternity wards using machine learning techniques. Comput Methods Programs Biomed. 2022;221.

16. World Bank. World Development Indicators [Internet]. 2025. Available from: https://databank.worldbank.org/source/world-development-indicators

17. Islam MM, Rahman MJ, Islam MM, Roy

DC, Ahmed NAMF, Hussain S, *et al.* Application of a machine learning based algorithm for prediction of malnutrition among women in Bangladesh. Int J Cogn Comput Eng. 2022;3:46–57.

18. Nejsum FM, Wiingreen, Freja Marie Nejsum 1 R, Jensen AK, Løkkegaard ECL, Hansen BM. Predicting early cessation of exclusive breastfeeding using machine learning techniques. PLoS One. 2025;20(1).

19. Yusuf JO, Miyim AM. Predicting breastfeeding practice of Nigerian child using machine learning and deep learning algorithms. Dutse J Pure Appl Sci. 2022;7(4a):183–93.

20. Talukder A, Ahammed B. Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition. 2020;78.

21. Zemariam AB, Abey W, Kassaw AK, Yimer A. Comparative analysis of machine learning algorithms for predicting diarrhea among under-five children in Ethiopia: Evidence from 2016 EDHS. Sage Journals Home. 2024;

22. Anku EK, Duah HO. Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms. PLoS One. 2024;19(2 February).