BIO Web of Conferences **243**, 01009 (2026) https://doi.org/10.1051/bioconf/202624301009
ICNF 2026

Exploring Predictors of Stunting Risk Through Machine Learning

Hardinsyah Hardinsyah[1*], *Rahmat* Oktavian[2], *Agus* Buono[2], *Salma* Widya Azhari[3], *Triagung* Yuliana[1], *Ova* Emilia[3], *Susaldi* Susaldi[3], *Yova* Agustini[3], *Meita* Dhamayanti[4], and *Fanni* Hanifa[4]

[1]Nutritionist Professional Education, Faculty of Medicine and Nutrition, IPB University, 16680 Bogor, Indonesia
[2]Computer Science Study Program, Faculty of Data Science, Mathematics and Informatics, 16680 Bogor, Indonesia
[3]Postgraduate Study Program in Nutrition Science, Faculty of Medicine and Nutrition, IPB University, 16680 Bogor, Indonesia
[4]Nursing Study Program, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia
[5]Department of Child Health, Faculty of Medicine, Bandung 40161, Indonesia

Abstract. Stunting has long-term health impacts, and short birth length is an early indicator of stunting. The application of Artificial Intelligence (AI) offers opportunities to build more accurate and interactive predictive models. This study aims to develop a classification model to predict the stunting status of children under 2 years old. Model for estimating stunting status by leveraging ML algorithms. The model was created using the XGBoost algorithm, implemented in Python with the scikit-learn module, and executed on Google Colab. Interactive input of attribute values allowed the system to predict lenght-for-age z-score (LAZ) categories. The XGBoost model built in this study was tested on a dataset of 72 samples. The variables showing the highest correlations, based on Spearman's rank correlation analysis, include the age of children under two, knowledge of complementary feeding (CF) practices, the presence of acute respiratory infection (ARI) symptoms, breastfeeding status, early initiation of breastfeeding (EIBF), maternal body weight, and the frequency of milk consumption. The XGBoost-based ML model achieved 90% accuracy in predicting stunting risk, demonstrating that the AI algorithm is reliable for data-driven stunting prediction.

1 Introduction

Child stunting remains one of the most pressing public health challenges of the twenty-first century, affecting approximately 149 million children under five years of age worldwide [1]. According to the World Health Organization (WHO), stunting is defined as impaired linear

*Corresponding author: hardinsyah@apps.ipb.ac.id

https://doi.org/10.1051/bioconf/202624301009

growth and development in children caused by chronic undernutrition, recurrent infections, and inadequate psychosocial stimulation. A child is classified as stunted when the Height-for-Age Z-score (HAZ) or Length-for-Age Z-score (LAZ) is below minus two standard deviations (< −2 SD) from the median of the WHO Child Growth Standards. HAZ is commonly used for children aged 24 months and older who are measured standing, whereas LAZ is used for children under 24 months who are measured in a recumbent position. Stunting reflects long-term nutritional deprivation and is associated with delayed cognitive development, poor educational performance, reduced productivity, and increased risk of morbidity and mortality [2, 3].

In the context of child malnutrition, machine learning approaches show particular promise for identifying at-risk children before clinical symptoms become apparent, potentially enabling earlier and more effective interventions. Among various machine learning algorithms, XGBoost (Extreme Gradient Boosting) has emerged as a particularly powerful tool for classification tasks in public health research. XGBoost is a gradient boosting library that provides parallel boosting trees algorithms capable of solving complex machine learning tasks [4]. Recent studies have demonstrated the exceptional predictive capabilities of ensemble methods for stunting classification, with XGBoost achieving 95.52% accuracy for stunting prediction in some contexts and up to 98.42% in others [5]. The algorithm's ability to handle non-linear relationships and complex interactions among socioeconomic, institutional, and dietary factors makes it especially suitable for nutrition outcomes research [1]. By leveraging data-driven approaches, health systems can move toward proactive nutritional surveillance, enabling early identification of vulnerable children at risk of stunting and facilitating evidence-based resource allocation [1]. This study addresses the critical need for early stunting detection by developing an AI model for predicting stunting status of children under two years old.

2 Method

This study employed a predictive modeling approach using machine learning techniques to classify stunting risk based on infant and maternal characteristics. The model was developed and validated using a dataset of 72 samples obtained from communities within the catchment areas of primary healthcare centers in Bogor Regency, West Java, Indonesia. The dataset included anthropometric measurements, health history variables, and nutritional intervention data obtained from previous studies, namely the *Gerakan Anak Sehat – Kolaborasi Inklusif Pengusaha Indonesia Atasi Stunting* (GAS KIPAS STUNTING), which was organized by the Association of Nutrition Higher Educations of Indonesia (AIPGI) and the Indonesian Employers Association (APINDO). The relatively small sample size reflects the tentative nature of this initial modeling effort, which serves as a proof-of-concept for applying XGBoost to stunting prediction [6, 7].

The predictors and features incorporated into the machine learning models were systematically structured into eight primary thematic domains, reflecting both immediate and underlying determinants of child stunting: (1) respondent characteristics (mothers & children under two years), (2) access & utilization of health services, (3) health status/morbidity, (4) supplementary feeding, (5) breastfeeding behavior, (6) breastfeeding practices, (7) pregnancy history and ANC services, and (8) mother's knowledge. The machine learning analyses were conducted at the individual-variable level, and only variables identified as important contributors are reported.

The dataset used consists of 72 child observations with various attributes representing demographic characteristics, health status, and children's food consumption patterns. The target variable in the dataset is represented by the JX column, which describes the growth status category based on the Length-for-Age Z-score (LAZ) indicator. For binary classification, the JX variable was grouped into two classes: stunting (label = 1) and non-stunting (label = 0). Categories indicating stunted and severely stunted conditions were combined as the stunting class, while normal and tall categories were combined as the non-stunting class. After the label transformation, the dataset consisted of 36 stunting and 36 non-stunting samples. Consequently, the dataset achieved a perfectly balanced class distribution, mitigating the class imbalance commonly encountered in epidemiological studies and public health surveys. This research develops a stunting prediction model using a machine learning approach based on Extreme Gradient Boosting (XGBoost), combined with Spearman correlation analysis, to identify attributes associated with stunting status. This approach allows analysis to be conducted in two stages: exploring statistical relationships among variables and building a predictive model to classify stunting status.

The model was developed in Python, leveraging core libraries such as pandas, NumPy, scikit-learn, XGBoost, SHAP, SciPy, and Matplotlib. A hold-out validation strategy was implemented, partitioning the dataset into training and test sets at an 80:20 ratio using the train_test_split function. This partitioning was performed using a stratified approach, provided that the test data sample size was sufficient to preserve the original class proportions. Model performance was comprehensively evaluated using a confusion matrix, accuracy, precision, recall, F1-score, and a classification report. To ensure the model's generalization ability, the dataset was split into training and test sets using a hold-out validation scheme with a 80:20 ratio. Model performance was evaluated using a confusion matrix on the test data to assess the model's ability to classify stunting status. These evaluation metrics provide an overview of the model's prediction accuracy and its ability to detect stunting cases accurately [6, 7].

This study used the XGBoost algorithm combined with the SHAP framework to identify predictors of stunting risk using a machine learning approach. Integrating bivariate Spearman correlation and multivariate SHAP analysis provides a robust framework for identifying and interpreting the primary drivers of stunting. While the initial Spearman correlation establishes a traditional baseline by identifying individual linear relationships (such as a child's age and maternal knowledge), the SHAP framework successfully captures the complex, non-linear dynamics and feature interactions driving stunting predictions [8]. The finding that child age has the strongest correlation with stunting status is epidemiologically consistent. Stunting is a cumulative process; it is not typically present at birth but develops over time due to chronic nutritional deficiencies and recurrent infections [9]. In alignment with the study's focus on using machine learning to explore predictive features, both methods consistently identify the chronological age of children under 2 years as the most dominant determinant (Figure 1 and Figure 2).

3 Results and discussion

The present study identified children's age as the strongest factor associated with stunting. The XGBoost model demonstrated strong predictive performance on the test dataset, achieving 90.38% accuracy, 91.35% precision, 90.38% recall, and an F1-score of 87.97%. These results indicate that the proposed model can accurately identify stunting status among children under 2 years old. Based on this satisfactory predictive performance, SHAP analysis was subsequently performed to interpret the contribution of each predictor, while Spearman rank correlation analysis was used to identify the variables most strongly associated with stunting. This finding is consistent with previous studies showing that the risk of stunting increases during the first two years of life, particularly between 6–24 months, which is considered the critical window for growth and development. Inadequate nutrient intake, recurrent infections, and inappropriate feeding practices during this vulnerable stage can lead to chronic growth faltering. Moreover, growth deficits occurring before the age of two years are often difficult to reverse later in life [10]. Mothers or caregivers with better knowledge regarding appropriate complementary feeding practices are more likely to provide timely, adequate, safe, and nutrient-dense foods for children. Poor caregiver knowledge causes inadequate feeding frequency, low dietary diversity, and insufficient nutrient intake, which contribute to impaired child growth and increased stunting risk [11]. Maternal body weight was also associated with stunting incidence in children. Maternal nutritional status reflects maternal energy and nutrient reserves before and during pregnancy, which strongly influence fetal growth and child development. Mothers with inadequate body weight are more likely to experience poor pregnancy outcomes, including low birth weight and intrauterine growth restriction, which increase the risk of stunting in later life [10].

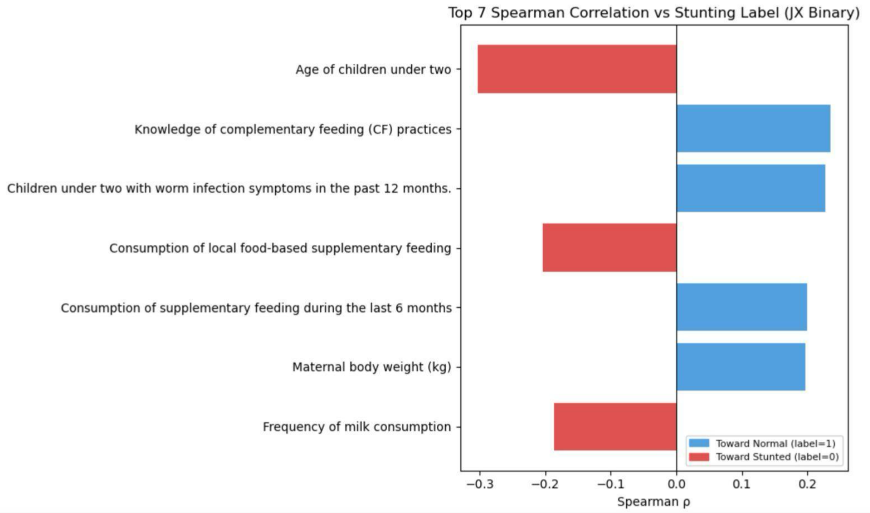

Fig 1. Seven Factors Associated with Stunting using Spearman Rank Correlation

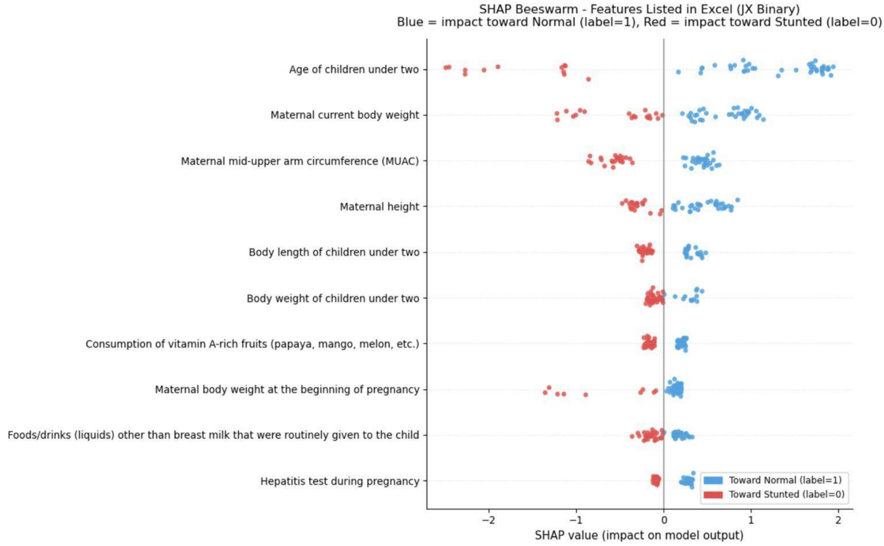

Fig 2. SHAP Beeswarm Plot of Global Feature Importance for Stunting Prediction

Children under two years with worm infection symptoms in the past 12 months also demonstrated a strong correlation with stunting. Helminth infections may impair nutrient absorption, reduce appetite, and increase nutrient losses, ultimately affecting children's nutritional status and linear growth. Recurrent infections additionally trigger chronic inflammation and weaken immunity, further exacerbating undernutrition [10]. Local food-based supplementary feeding programs can improve nutrient intake by using locally available, affordable, and culturally acceptable foods rich in protein and micronutrients. Previous studies reported that supplementary feeding with locally available ingredients improved body weight and height among children at risk of stunting. The use of local foods may also enhance sustainability and community participation in nutrition programs while supporting food security at the household level [11].

Strong multicollinearity was observed, particularly among nutritional intervention variables, which may affect the interpretation of feature importance in the machine learning model. Although XGBoost is relatively robust to multicollinearity, highly correlated variables can cause feature importance scores to become unstable and distributed across related predictors. This finding suggests that the identified nutritional factors likely represent similar aspects of the child's nutritional support environment and should therefore be interpreted with caution. The performance of machine learning algorithms is highly influenced by the structure and quality of the dataset, feature selection, and the characteristics of the problem being analyzed. Therefore, the results obtained are context-specific and cannot be universally generalized. In addition, model performance depends on the quality of the available data and the system capacity used during model development. Although machine learning approaches can help identify predictors of stunting risk, developing predictive models alone does not directly reduce the prevalence of stunting. Comprehensive and sustainable intervention efforts are still required to address this issue

4 Conclusion

This study successfully developed a tentative machine learning model using the XGBoost algorithm to estimate the risk of stunting among children under 2 years old, achieving an accuracy of 90%. The model identified seven important factors associated with stunting, namely age of children, knowledge of complementary feeding practices, worm infection symptoms in the past 12 months, consumption of local food-based supplementary feeding, consumption of supplementary feeding during the last six months, maternal body weight, and frequency of milk consumption. However, the performance of machine learning algorithms is highly dependent on dataset quality, feature selection, and the characteristics of the problem being analyzed. Although machine learning can help identify predictors of stunting risk, comprehensive and sustainable intervention efforts are still required to reduce stunting prevalence.

The authors would like to acknowledge the financial support provided by the DAPT EQUITY Program under the Indonesian Collaborative Research Scheme A, under Contract Number 60374/IT3/PT.01.03/P/B/2025. The support from this program greatly contributed to the implementation and completion of this research.

References

1. W.R. Mgomezulu, P. Thangata, B. Mkandawire, N. Amoah, Advancing predictive analytics in child malnutrition: Machine, ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting. Hum. Nutr. Metab. **42,** 200340 (2025). https://doi.org/10.1016/j.hnm.2025.200340.

2. WHO, *Global Nutrition Targets 2030 : Stunting Brief*. Geneva, Switzerland (2025).
3. WHO, *Malnutrition in Children*. Available at: https://www.who.int/data/nutrition/nlis/info/malnutrition-in-children (Accessed Apr. 30, 2026).
4. Z.L.R. Kim, S. Vollmer, S.V. Subramanian, Factors associated with child stunting, wasting, and underweight in 35 low- and middle-income countries. JAMA Netw. Open **3**, e203386 (2020). https://doi.org/10.1001/jamanetworkopen.2020.3386.
5. F. La Rochelle, Performance of Random Forest and XGBoost in stunting classification for toddlers in Talaud Island. 2025 Int. Conf. Inf. Commun. Technol. 1–6 (2025). https://doi.org/10.1109/ICoICT66265.2025.11193038.
6. H. Shen, H. Zhao, Y. Jiang, Machine learning algorithms for predicting stunting among under-five children in Papua New Guinea. Children. **10**, 1638 (2023). https://doi.org/10.3390/children10101638.
7. E.B. Weiser, Leveraging artificial intelligence in public health: A comparative evaluation of machine-learning algorithms in predicting COVID-19 mortality. J. Stud. Internet Med. **8**, 1–18 (2025). https://doi.org/10.29024/jsim.216.
8. S.M. Lundberg, S. Lee, A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. (NeurIPS) 30 (2017). https://doi.org/10.48550/arXiv.1705.07874.
9. C.G. Victora, P. Christian, L.P. Vidaletti, G. Gatica-Domínguez, P. Menon, R.E. Black, Revisiting maternal and child undernutrition in low-income and middle-income countries: Variable progress towards an unfinished agenda. Lancet 397, 1388–1399 (2021). https://doi.org/10.1016/S0140-6736(21)00394-9.
10. S. Suyitno, M. Maretalinia, M.U. Khan, S. Supriatin, Appropriate complementary feeding among Indonesian children aged 6-23 months: A descriptif literature review, Ahmar Metastasis Health J., **4,** 2, 117–126 (2024). https://doi.org/10.53770/amhj.v4i2.317.
11. D. Ghita, Complementary feeding and stunting: A literature review. Hasanuddin Med. Stud. J. **6**, 2 (2025). https://doi.org/10.24252/hmsj.v6i2.57864.