Stunting Prediction in South Sumatra Province, Indonesia: Machine Learning Approaches Using SKI 2023

Alfensi Faruk[1], Dian Cahyawati[2], Endang Sri Kresnawati[3]
{alfensifaruk@unsri.ac.id[1], dianc_mipa@unsri.ac.id[2], eskresna@unsri.ac.id[3]}

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sriwijaya[1,2,3]
Corresponding author: alfensifaruk@unsri.ac.id

Abstract. Stunting is still one of the main health issues in Indonesia, including in South Sumatra Province. This study utilises logistic regression and gradient boosting machines using the Indonesian Health Survey (SKI) 2023 under-five children microdata ($n = 2,940$). The stunting prevalence in South Sumatra had decreased from 20.3% in 2023 to 15.9% in 2024. However, more effort is still required to achieve the national target. Our results show that birth weight, birth length, living conditions, sanitation, and the child's sex are factors that significantly affect stunting, confirmed also by SHapley Additive exPlanations (SHAP) analysis. Both models show modest discrimination capability (AUROC $\approx$ 0.59). Our findings are informative to enhance the strategies of stunting prediction.

Keywords: Stunting Prediction, under-five children, national survey data, logistic regression, gradient boosting machine

1 Introduction

Stunting occurs when a child's growth and development do not follow standard references for their age [1]. The global prevalence was about 23% among children under five, and the rates were much higher in Asia and Africa, where people are more likely to be vulnerable. Stunting is a sign of long-term undernutrition, and it also keeps going in families, causing poor health, cognitive impairment, lower educational achievement, and lower economic productivity. The United Nations Sustainable Development Goals (SDGs) set very high goals to achieve zero hunger, make sure everyone has enough food, and get rid of all forms of malnutrition (including stunting) by 2030.

Stunting is still one of the main public health issues in Indonesia. Despite the national stunting prevalence for under-five children decreasing from 21.5% in 2023 to 19.8% in 2024, it is still far from the 2029 target (14%). According to the Indonesian Health Survey (SKI) 2023 [2], the prevalence of stunting in the South Sumatra Province was recorded at 20.3% in the year 2023. Importantly, South Sumatra has made notable advancements, decreasing its stunting prevalence to 15.9% in 2024 and

SICBAS 2025, November 06, Palembang, Indonesia

DOI 10.4108/eai.6-11-2025.2364517

being recognized as the second most effective province in Indonesia concerning initiatives aimed at alleviating stunting. However, notwithstanding these gains, South Sumatra must maintain and intensify its actions. The gap to the national target remains (14% in 2029), and sustained reductions are vital to maintaining the health, cognition, and long-term productivity of its children.

Numerous studies demonstrated that maternal health status, educational attainment, nutritional standards, unfavorable socioeconomic conditions, availability of healthcare services, and inadequate access to potable water and sanitation constitute pivotal determinants influencing stunting [3, 4, 5]. However, individual-level prediction accuracy and more effective interventions are still challenging. In practice, a number of standard machine learning approaches, such as gradient boosting machines (GBM) and logistic regression (LR), also gain more attention. LR has an advantage in interpretability, and it is capable of determining relevant factors impacting stunting. Meanwhile, the gradient boosting machine can capture nonlinear patterns in the data.

The primary aims of this investigation are to formulate and assess predictive models for stunting in children under the age of five within the South Sumatra Province, utilizing LR alongside the GBM algorithm. In particular, we seek to (1) ascertain the principal determinants that correlate with the risk of stunting, (2) evaluate and compare the predictive accuracy, calibration, and discriminatory efficacy of both logistic regression and the gradient boosting machine, and (3) furnish empirical evidence for the implementation of targeted, region-specific interventions aimed at expediting the reduction of stunting in accordance with national priorities.

2 Methods

2.1 Data preprocessing and covariates

In this paper, we employed microdata from the SKI 2023 [2] at the individual level for children under-five years of age in South Sumatra Province, Indonesia. The initial sample consists of $n = 2,979$ children, where the main outcome variable was stunting (*stunted*). To compute each child's stunting status, we applied the WHO guidelines [6]. If a child's height for age z-score (*HAZ*) is below -2 standard deviations, he/she was categorized as stunted. The *HAZ* formula [7] used is

$$HAZ = \begin{cases} \frac{\ln(H/M)}{S} & \text{if } L = 0, \\ \frac{\left(\frac{H}{M}\right)^L - 1}{L \times S} & \text{if } L \neq 0, \end{cases} \tag{1}$$

where H is observed height/length of the child, M is median height from the reference population, L is Box-Cox power to account for skewness, and S is coefficient of variation to account for dispersion.

We grouped the potential covariates by measurement scale (continuous or categorical). We dropped rows with a missing *stunted* as the target, and converted the target variable to an integer for modeling. To handle missing values in the covariates, we conducted three strategies: (1) low missing ($< 5\%$), (2) medium missing (5% to 30%), and (3) high missing ($> 30\%$). For low missing, we replaced the missing values of continuous variables using the median method, and the mode approach for categorical variables. For medium missing, the missing values of continuous variables were filled by modeling each variable as a function of other variables. Meanwhile, for the categorical variables, we filled the missing values using mode approach. For the high missing, we dropped the variables except some forced categorical variables, such as *mr_edu*, *mr_job*, *drwater_source*, and *toilet*. The latter were still imputed using mode approach. After handling the missing values, the number of data slightly decreased to $n = 2,940$ under-five children. The potential determinants of the stunting of this work are given in Table 1.

To cope with the complex survey design, we applied *w_final* for weighting the label. Next, we randomly split the data into training (80%) and testing (20%). To prevent children within the same primary sampling unit from appearing in both the training and testing data (data leakage), the splitting within cluster was defined by variable *psu*.

Table 1: Respondents' characteristics

Covariate	Description	Measurement Scale
age_month [8, 9]	Child's age in months	Continuous
birth_weight [10, 11]	Weight at birth (grams)	Continuous
birth_length [12, 11]	Length at birth (cm)	Continuous
house_area [13]	Dwelling size (square meters)	Continuous
sex [10, 13]	Child's sex	Nominal
mr_edu [10]	Mother's education level	Ordinal
mr_job	Mother's employment status	Nominal
urban [13]	Place of residence	Nominal
drwater_improved [13]	Improved drinking water	Nominal
toilet_improved [13]	Improved sanitation facility	Nominal

2.1.1 Logistic regression

Logistic regression (LR) or Logit model is a standard machine learning approach for predicting the probability of outcomes based on one or more covariates. We apply LR to stunting prediction, and we group the label into two classes, i.e. stunted (category 1) and not stunted (category 2).

We use LR to predict the probability of stunting (p) based on a set of covariates ($x_1, x_2, \cdots, x_m$). The formula of LR is written as follows

$$\ln\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m, \tag{2}$$

where p is probability of category 1, β_0 is intercept and β_i is regression coefficients for the i^{th} covariate.

By transformation of (2), the probability of stunting is

$$p = \frac{1}{1 + \mathrm{e}^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}}. \tag{3}$$

We are also interested in the odds ratio (OR), defined as

$$\mathrm{OR} = e^{\beta_j}, \tag{4}$$

where $j = 1, 2, \cdots, m$ showing how much the odds of stunting change with a one-unit increase in covariate x_j. The loss function of LR is

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln(p_i) + (1-y_i)\ln(1-p_i)\right], \tag{5}$$

where N is the number of data, y_i is true value of label for the i^{th} individual, and p_i stunting probability for the i^{th} individual.

2.1.2 Gradient boosting machines

Gradient Boosting Machines (GBM) is one of ensembles approaches combining a number of decision trees. The m^{th} iteration of GBM is defined as follows

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \tag{6}$$

where $F_m(x)$ is model in the m^{th} iteration, γ_m= learning rate in the m^{th} iteration, and $h_m(x)$ is the new estimator to fix the residual error in the $(m-1)^{th}$ iteration.

The objective of GBM is to predict the values of $y_i = F_m(x_i)$ by correcting the error of $F_{m-1}(x_i)$. Thus, we fit the estimator $h_m(x_i)$ to the residual $y_i - F_m(x_i)$. In general, GBM minimises

$$L = \sum_{i=1}^{N} l(y_i, F(X_i)), \tag{7}$$

where $l(y_i, F(X_i))$ is the loss function between the true value y_i and the predicted value $F(x_i)$ for the i^{th} individual, and N is the number of individuals.

3 Results

3.1 Sample characteristics and prevalence

The cohort analyzed in the present study consists of 2,940 children under the age of five residing in South Sumatra. As illustrated in Table 2, the weighted prevalence of stunting is determined to be 19.3%, which is in close alignment with the national stunting prevalence reported in 2023 (21.5%).

The descriptive statistics, as evidenced in Table 2, reveal that children characterized by stunting were, on average, older (33.49 months in comparison to 31.94 months). The children with lower birth weight and lower birth length were prone to being stunted. Larger house areas were also the living environment of most of the non-stunted children. In categorical variables, the table shows the proportion (in %) of each category. The male children had a higher risk of stunting (55%) than the female (45%). Stunting prevalence was also greater in rural areas (62%) than in urban areas

(38%). Furthermore, children whose mothers had little education were more vulnerable to stunting. The table also suggests a higher prevalence for children who were living with unimproved toilet facilities (89%) compared to those with such facilities (11%). In general, these patterns align with socio-environmental gradients identified in national surveys.

Table 2: Summary statistics by stunting status.

Variable	Categories	Overall	Non-stunted	Stunted
age_month (months)	-	32.24	31.94	33.49
birth_weight (grams)	-	3084.71	3110.53	2976.99
birth_length (cm)	-	48.44	48.63	47.66
house_area (m^2)	-	72.57	74.62	64.04
sex (0)	Male	0.51	0.50	0.55
sex (1)	Female	0.49	0.50	0.45
mother_edu (0)	No Education	0.02	0.01	0.03
mother_edu (1)	Primary	0.20	0.19	0.22
mother_edu (2)	Secondary+	0.78	0.79	0.75
urban (0)	Urban	0.40	0.40	0.38
urban (1)	Rural	0.60	0.60	0.62
mother_job (0)	No	0.73	0.73	0.74
mother_job (1)	Yes	0.27	0.27	0.26
drwater_improved (0)	Yes	0.05	0.05	0.05
drwater_improved (1)	No	0.95	0.95	0.95
toilet_improved (0)	Yes	0.07	0.07	0.11
toilet_improved (1)	No	0.93	0.93	0.89
stunted (0)	No	0.81	1.00	0.00
stunted (1)	Yes	0.19	0.00	1.00

Note: Continuous variables are presented as means. Categorical variables are presented as proportions.

3.2 Selected determinants

In this study, we explored two strategies to determine the main risk factors for stunting in children: multivariable logistic regression (using backward stepwise selection) and SHAP value analysis. For regression, all candidate covariates were initially included in a full model. Covariates with a p-value above 0.05 were considered for removal, applying the process step by step and monitoring the Akaike Information Criterion (AIC) to assess model fit. When removing a variable failed to improve the AIC, the selection process was stopped, and the final set of covariates was established.

Variable selection resulted in no meaningful AIC change. The full model (AIC=538258.45) and its first reduced version (AIC = 538260.86) were nearly identical. Hence, all covariates were retained. Based on these results, presented in Table 3, birth weight, birth length, house area, sex, and toilet facility status emerged as significant covariates for childhood stunting. While other covariates, such as urban residence, mother's occupation, drinking water, and maternal education showed weaker statistical evidence, they were kept in the model to ensure comprehensive adjustment for potential confounders. Overall, the most obvious correlations with lower stunting risk are found in household living conditions, sanitation, and growth at birth (weight/length); the other sociodemographic factors in this model do not exhibit any independent effects.

The logistic regression suggests a few risk elevations and a number of protective factors against stunting. Lower odds of stunting are indicated by statistically significant predictors (5% level) with odds ratios (*OR*) less than one: more birth weight is protective ($OR = 0.999$ per gramme, $p < 0.001$), which is more interpretable as approximately 7% to 8% lower odds per 100 g increase; greater birth length is strongly protective ($OR = 0.935$ per cm, $p < 0.001$; about 6% to 7% lower odds per additional cm); larger house area is protective ($OR = 0.994$ per unit, $p = 0.0028$; about 5% lower odds per greater than 10 units); and having an improved toilet is linked to approximately 41% lower odds ($OR = 0.588$, $p = 0.016$).

Table 3: Logistic regression results for determinants of stunting.

Covariate	*OR*	95% CI (Low)	95% CI (High)	p-value	Sig. (5%)
age_month	1.0048	0.9971	1.0125	0.2268	No
sex	0.6930	0.5260	0.9131	0.0092	Yes
birth_weight	0.9992	0.9989	0.9996	1.78×10^{-5}	Yes
birth_length	0.9348	0.9048	0.9658	5.07×10^{-5}	Yes
mother_edu	0.7662	0.5835	1.0061	0.0553	No
mother_job	0.9837	0.7337	1.3188	0.9125	No
house_area	0.9944	0.9907	0.9981	0.0028	Yes
urban	0.9665	0.7135	1.3091	0.8256	No
drwater_improved	0.9005	0.5172	1.5680	0.7112	No
toilet_improved	0.5884	0.3821	0.9058	0.0160	Yes

3.3 SHAP analysis

We also employed SHapley Additive exPlanations (SHAP) value analysis on GBM using the tuned hyperparameters from 5-fold cross validation, i.e. 'learning_rate': 0.02, 'max_depth': 2, 'max_features': 'log2', 'min_samples_leaf': 2, 'min_samples_split': 6, 'n_est': 324, 'subsample': 0.89. This SHAP procedure aims to demonstrate an overall view of variable importance and the possible nonlinear relationships in the model. Meanwhile, the backward stepwise selection method in the logistic regression captures any statistically significant associations of each covariate with the stunting status.

Figure 1 visualizes the SHAP analysis from the tuned GBM for the South Sumatra stunting data. Each individual child represented in the dataset is symbolized by a dot within the graphical representation. The horizontal axis delineates the extent to which each covariate influences the predicted probability of stunting, with positive values denoting an elevated risk and negative values suggesting a diminished risk. Red points signify elevated feature values, while blue points indicate reduced feature values. The SHAP analysis shows that house area, birth weight, sex, age, birth length, and mother's education are the most influential covariates of under-five children's stunting status in South Sumatra Province. Male children who live in smaller houses, with low birth weight and length, had a higher risk of stunting. These results are in line with the results from the backward stepwise selection in logistic regression. This SHAP analysis moreover demonstrates nonlinear influences of covariates on children's stunting. The SHAP analysis shows that house area, birth weight, sex, age, birth length, and mother's education are the most influential covariates of under-five children's stunting status in South Sumatra Province. Male children who live in smaller houses, with low birth weight and length, had a higher risk of stunting. These results are in line with the results from the backward stepwise selection in logistic regression. This SHAP analysis moreover demonstrates nonlinear influences of covariates on children's stunting.

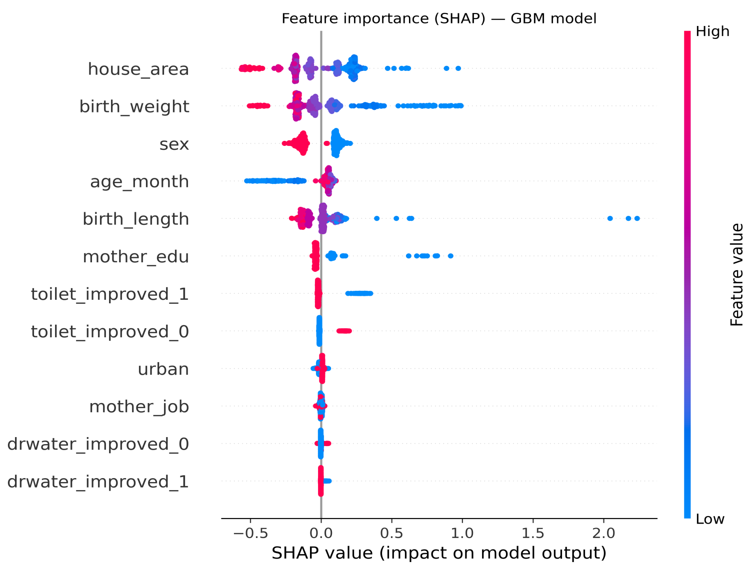


Fig. 1. SHAP summary from GBM for South Sumatra data.

3.4 Models' predictive performance

In the preprocessed stage, we have randomly divided the original data into training data (80%) and testing data (20%). The training data were used to train logistic regression (LR) and GBM. After the training process was done, we conducted prediction and model predictive performance evaluation on the independent testing data. The model's performance across five metrics is shown in Table 4. In general, the presented metrics can be grouped into calibration (Accuracy, F1 Score, and Brier score) and discrimination, namely Area Under the Receiver Operating Characteristic curve

(AUROC) and Area Under the Precision-Recall Curve (AUPRC). Calibration and discrimination assess the model's performance on two different aspects. Thus, they are incomparable. Calibration evaluates how close the outcomes match the outputs. Meanwhile, discrimination measures how well the models separate each individual into two groups.

Table 4: Performance metrics of predictive model.

Model	Accuracy	F1 Score	Brier Score	AUROC	AUPRC
LR	0.795	0.008	0.162	0.592	0.243
GBM	0.801	0.029	0.162	0.581	0.249

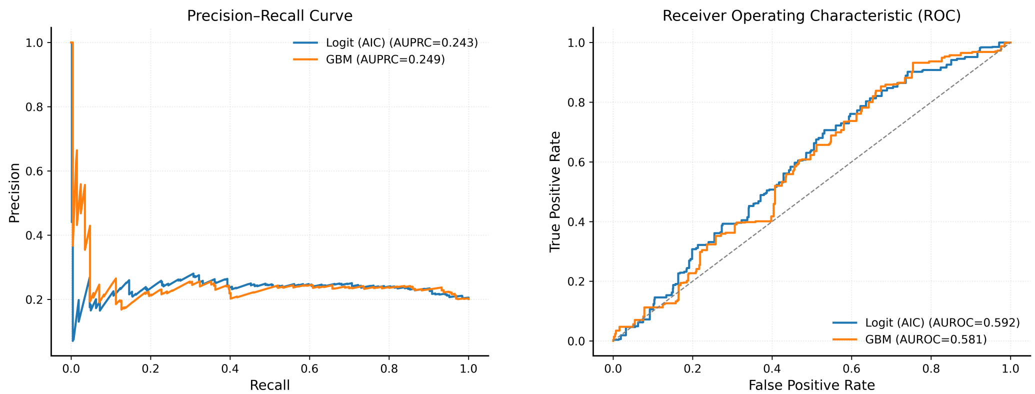
From Table 4, although both models' accuracy was around 0.80, the extremely low F1 scores (0.01 for LR and 0.03 for GBM) suggest that the classifiers miss positives and mostly predict the majority class, which is misleading given the stark class imbalance (19% stunting class). The probability estimates may need post-hoc calibration because the calibration was also suboptimal (Brier score ≈ 0.16 for both models), which is worse than a perfectly calibrated constant predictor at the base rate ($p(1-p) \approx 0.154$).

The tuned GBM had an AUROC of 0.581 and AUPRC of 0.249, which was slightly higher than the AUROC of 0.592 and the AUPRC of 0.243 for LR (Table 4). These numbers reveal that the models are not very strong at telling the difference between groups (AUROC = 0.5), but they still pick up key signals because there are not many socioeconomic and household factors. The Precision-Recall (PR) curves demonstrate that all of the models are more accurate than the baseline (≈ 0.19, which is the same as the rate of stunting class). This means that there is a big improvement in prediction, even while the overall recall is poor. The Receiver Operating Characteristic (ROC) and PR trends for LR and GBM models are essentially the same. This illustrates that both traditional and ensemble methods reveal risk gradients that are similar. This is in accordance with the shared factors reported in regression and SHAP analysis in the previous subsections.

As the complement of Table 4, we also amend the plots of AUCROC and AUPRC as presented by Figure 2. The figure shows the PR curves closely resemble the no-skill baseline for most levels of recall. From Panel 2a, there are only small spikes at near-zero recall (when a few high-precision predictions happen), and then they settle around precision $0.20-0.25$ across moderate recall, which is in accordance with the low F1 values. The ROC curves (Panel 2b) are just a bit higher than the diagonal, and for any given false-positive rate, the true-positive rate rises, where LR is slightly higher than GBM. This means that the classes do not really stand out from each other. GBM is better than LR on PR-oriented measures (higher AUPRC and F1), while logistic regression is better on ROC (AUROC 0.59 vs 0.58); the differences are modest and neither model is strong.

4 Discussion

In this work, we have revealed several factors affecting the stunting in South Sumatra. An elevated birth weight, increased birth length, and improved household sanitation (toilet facilities)



(a) **(b)**

Fig. 2. Model discrimination performance: (a) PR curve, and (b) ROC curve.

were associated with a significantly reduced incidence of stunting, corroborating established biological and environmental determinants [14, 15]. From the obtained odds ratio for the sex variable ($OR = 0.693, p = 0.009$), females experience a diminished risk in comparison to their male counterparts. This result is consistent with current literature [16, 17, 18]. Socioeconomic status, as indicated by residential area, exhibited a minimal effect on the prevalence of stunting. In other words, there is an indication that stunting might be affected by household wealth [18, 19].

When we checked to see how well the model worked, we found that it could not predict well. The AIC-selected logistic model and GBM both had accuracies of about 0.80, but their F1-scores were not very high ($0.01-0.03$). This means that there was a lot of class imbalance, and they were not very good at finding stunted cases. We also found that the discrimination capability of the fitted models was slightly above random guessing, either based on AUROC or AUPRC (see [20] for more details). Brier scores also showed that the calibration of the model was like a simple constant model. The results indicated that the used covariates and the chosen models do not sufficiently conduct more accurate prediction [21]. Based on F1 and AUPRC, it is shown the advantage of GBM in capturing nonlinear relationships despite the current benefits not being significant enough [20]. The potential cause of the less accurate prediction was that the data used did not include a lot of variables, had errors in measurement, and did not include some critical factors, such as dietary habits and nutrition of the mother [22].

5 Conclusion

This research demonstrates that although significant biological and environmental determinants, notably birth outcomes and sanitation, persist as robust correlates of stunting in South Sumatra, their collective predictive efficacy for identifying at-risk children remains constrained. The LR and GBM models are only slightly superior to random chance classification, highlighting the imperative for improved data integrity and advanced model refinement. Future work should include additional potential independent variables, such as dietary habits and nutrition of the mother. These variables might be obtained from other datasets, namely the Indonesian Demographic and Health Survey (IDHS). Applying more advanced ML approaches with class balance and studying regional variation would also potential direction as the future study.

References

[1] World Health Organization. Stunting in a nutshell. Geneva: World Health Organization; 2015. [cited 2025 Oct 25]. World Health Organization [Internet]. Available from: https://www.who.int/news/item/19-11-2015-stunting-in-a-nutshell.

[2] Badan Kebijakan Pembangunan Kesehatan KKRI. Survei Kesehatan Indonesia (SKI) 2023. Jakarta: Kementerian Kesehatan Republik Indonesia; 2024. Available from: https://badankebijakan.kemkes.go.id.

[3] Saleh A, Syahrul S, Hadju V, Andriani I, Restika I. Role of maternal in preventing stunting: a systematic review. Gaceta sanitaria. 2021;35:S576-82.

[4] Arini HRB, Peranto S. Social analysis of childhood stunting in Indonesia. The Southeast Asian Journal of Tropical Medicine and Public Health. 2023;54(Suppl 1):21-38.

[5] Handayani NS, Huriyati E, Hasanbasri M. Association of maternal education with nutritional outcomes of poor children with stunting in Indonesia. Asia Pacific Journal of Public Health. 2023;35(5):373-80.

[6] Bloem M. The 2006 WHO child growth standards. British Medical Journal Publishing Group; 2007.

[7] Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Statistics in medicine. 1992;11(10):1305-19.

[8] Trinh MH, Choi Y, Liu C, Finucane ME, Smyth JD, Baker DA, et al. Trajectory and Covariates of Children's Screen Media Time. JAMA Pediatrics. 2019;173(4):385-91.

[9] Thompson KM, Eisenberg DM, Wolpert D. Changes in Children's Exposure as a Function of Age and the Relevancy of Control Variables: A Methodological and Empirical Investigation of Age Effects. Environmental Health Perspectives. 2004;112(3):329-35.

[10] Wubetu AD, Zerfu AA, Muche HA. Newborn Birth Weight and Associated Factors Among Mother-Infant Pairs in Public Hospitals in Ethiopia: A Cross-Sectional Study. PLoS ONE. 2021;16(3):e0249021.

[11] Yokoyama Y, Sawada A, Miyake D, Sasaki S. Genetic and Environmental Factors Affecting Birth Size Variation: A Twin Study. International Journal of Epidemiology. 2018;47(4):1195-206.

[12] Krebs NF, Alder SC, Baenziger JJS. Birth Length is the Strongest Predictor of Linear Growth Status and Stunting at Two Years: A Multi-center Analysis. Frontiers in Nutrition. 2022;9:935413.

[13] Knapp EA, Gottfredson LL, Zhuang D, Farr SL, Rappaport SN. Environmental Influences on Child Health Outcomes (ECHO): Study Protocol and Key Variables. American Journal of Epidemiology. 2023;192(8):1249-60.

[14] Dwivayana IKBK, Setiawan KH, Agustini NNM, Pasek MS. The Relationship Between Birth Weight and Stunting in Toddlers at Kintamani VI Community Health Center in 2025. Ganesha Medicina. 2025;5(2):100-7.

[15] Ridha A, Maarif MZ, Noviati TD, Renowening Y, Mahmudah H, Fitriani RJ. Review of Factors Influencing the Impact of Access to Sanitation Facilities on Stunting. Jurnal eduhealth. 2023;14(02):701-6.

[16] Thurstans S, Opondo C, Seal A, Wells J, Khara T, Dolan C, et al. Boys are more likely to be undernourished than girls: a systematic review and meta-analysis of sex differences in undernutrition. BMJ global health. 2020;5(12):e004030.

[17] Wamani H, Åstrøm AN, Peterson S, Tumwine JK, Tylleskär T. Boys are more stunted than girls in sub-Saharan Africa: a meta-analysis of 16 demographic and health surveys. BMC pediatrics. 2007;7(1):17.

[18] Beal T, Tumilowicz A, Sutrisna A, Izwardy D, Neufeld LM. A review of child stunting determinants in Indonesia. Maternal & child nutrition. 2018;14(4):e12617.

[19] Rizal MF, Van Doorslaer E. Explaining the fall of socioeconomic inequality in childhood stunting in Indonesia. SSM-population health. 2019;9:100469.

[20] Adhikari S, Normand SL, Bloom J, Shahian D, Rose S. Revisiting performance metrics for prediction with rare outcomes. Statistical methods in medical research. 2021;30(10):2352-66.

[21] Farooq V, Brugaletta S, Vranckx P, Serruys PW. A guide to interpreting and assessing the performance of prediction models. EuroIntervention. 2011;6(8):909-12.

[22] Ndagijimana S, Kabano IH, Masabo E, Ntaganda JM. Prediction of stunting among under-5 children in Rwanda using machine learning techniques. Journal of preventive medicine and public health. 2023;56(1):41.