Z-Score-Free Stunting Prediction from Basic Anthropometric Measurements Using Machine Learning

[1]Muhammad Resha, [2]Apriana Toding
[1]Information System, Universitas Teknologi Akba Makassar, Makassar,
[2]Electrical Engineering, Universitas Kristen Indonesia Paulus, Makassar
[1]mresha@akba.ac.id, [2]apriana.toding@ukipaulus.ac.id

Early detection of stunting commonly relies on the WHO height-for-age Z-Score, which requires reference standards, calculation tools, and correct interpretation. These requirements can reduce screening efficiency in resource-limited community health services. This study proposes and validates a Z-Score-free machine learning framework for toddler stunting screening using only basic anthropometric measurements: age, sex, weight, and height. The scientific novelty lies in the deliberate exclusion of Z-Score-derived variables, weight-for-age status, weight-for-height status, and other nutritional-status indicators from the predictor set to prevent data leakage. Four yearly datasets from Jeneponto Regency, Indonesia, were combined, producing 40,071 labeled toddler records after removing one blank row. SMOTE was applied only to training data, and missing numeric values were handled through median imputation inside the training pipeline. Four algorithms were evaluated: XGBoost, Random Forest, LightGBM, and Logistic Regression. Under the valid Z-Score-free pipeline, Random Forest achieved the highest hold-out accuracy and F1-score (95.81% accuracy, 94.86% F1-score), while LightGBM achieved the highest ROC-AUC (99.28%) and recall (95.83%). XGBoost remained highly competitive (95.00% accuracy, 93.90% F1-score, and 99.12% ROC-AUC), but it did not outperform all competing classifiers on this dataset. Stratified 5-fold cross-validation confirmed the same overall pattern, with Random Forest producing the highest mean F1-score. SHAP analysis showed that height and age were the dominant contributors to the XGBoost decision process, aligning with the biological basis of stunting as impaired linear growth relative to age. The proposed framework provides a lightweight and scientifically more valid screening approach for Posyandu implementation because it avoids using deterministic diagnostic variables as model inputs.

Keywords — Stunting, Machine Learning, Z-Score, Anthropometric, Random Forest, XGBoost

I. Introduction

Stunting remains a major public health problem among children under five, particularly in low- and middle-income settings. It reflects chronic linear growth failure caused by prolonged nutritional deficiency, repeated infection, and inadequate environmental or caregiving conditions. The long-term consequences include impaired physical growth, delayed cognitive development, lower school performance, reduced productivity, and poorer health outcomes in adulthood [1]. The WHO defines stunting using the height-for-age indicator, where a child is categorized as stunted when the height-for-age Z-Score is below minus two standard deviations from the median of the WHO Child Growth Standards [2]. Although this standard is widely used and accepted in clinical practice, its practical application is dependent on accurate measurement, reference tables, digital calculators, or manual analysis of growth curves[3]. These prerequisites can reduce the effectiveness of screening and increase the risk of delayed detection, especially in community-based health services such as Posyandu in resource-limited areas.

Recent advances in machine learning have opened new possibilities for automating nutritional status assessments and for supporting early detection of stunting[4]. Several previous studies have applied classification algorithms such as Logistic Regression, Support Vector Machine, Random Forest, XGBoost, LightGBM, and hybrid machine learning frameworks to predict stunting or nutritional status using anthropometric, demographic, socioeconomic, and environmental variables[5]. These studies show that machine learning can improve prediction performance and identify important risk factors. However, much of the existing literature still focuses mainly on model accuracy, while the methodological validity of the input features receives less attention. In particular, some prediction models include Z-Score values, nutritional status labels, or other derived anthropometric indicators as part of the feature set. This practice can create data leakage because the model is trained using variables that are mathematically close to, or directly derived from, the target definition itself. This limitation is important because the inclusion of Z-Score-derived variables may cause machine learning models to behave merely as automated translators of the WHO deterministic formula rather than as independent predictive systems. As a result, the reported performance may be overestimated and may not represent the true ability of the algorithm to learn biological growth patterns from basic field measurements[6]. In practical deployment, this also weakens the value of the model because field cadres still need access to Z-Score calculation tools before the prediction can be made. Therefore, an important research gap remains: there is limited empirical evidence on whether high-performance machine learning models can accurately classify stunting using only fundamental anthropometric variables, without relying on Z-Score or other derived diagnostic features[7].

A clear research gap therefore remains: limited empirical evidence shows whether high-performance stunting classification can be achieved using only fundamental

anthropometric measurements while excluding all Z-Score-derived and diagnostic status variables from the feature set[8]. Another gap concerns validation. Many studies report only a single train-test split, while medical classification tasks require more robust evidence through stratified cross-validation, extended metrics, and statistical comparison among models[9].

Another gap concerns data context and local generalizability. Many previous studies have used public datasets, national survey datasets, or relatively small samples, which may not fully represent local growth patterns in specific high-risk regions of Indonesia. Stunting risk is strongly influenced by local demographic, nutritional, and environmental conditions; therefore, models trained on local primary health records may provide more relevant evidence for regional decision-support systems[10]. In addition, few studies have explicitly compared linear models with tree-based ensemble models under a controlled feature-isolation strategy, where all Z-Score-related variables are removed before training. Such comparison is necessary to determine whether non-linear ensemble architectures can extract sufficient predictive information from raw anthropometric measurements alone[11].

To address these gaps, this study proposes a data-centric machine learning approach for independent stunting prediction using only basic anthropometric variables, namely age, sex, weight, and height. A large-scale dataset consisting of 40,071 toddler health records from Jeneponto Regency, Indonesia, collected between 2021 and 2024, was used as the empirical basis of the study. All Z-Score columns and other derived nutritional indicators were removed during preprocessing to prevent data leakage and to ensure that the model learned directly from raw measurement patterns. To handle class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied only to the training data. Four algorithms, namely XGBoost, Random Forest, LightGBM, and Logistic Regression, were then benchmarked to evaluate the comparative performance of non-linear ensemble models against a conventional linear baseline.

The novelty of this study lies in three main aspects. **it introduces an independent Z-Score-free prediction framework that deliberately excludes deterministic diagnostic variables from the feature set.** Second, **it evaluates the capability of tree-based ensemble learning to capture non-linear relationships among fundamental anthropometric variables in a large local Indonesian toddler dataset**. Third, **it provides interpretable evidence through feature importance analysis to examine whether the model's decision pattern aligns with clinical growth principles**. The findings are expected to provide a scientific foundation for developing lightweight, low-cost, and real-time stunting decision-support systems that can be deployed in Posyandu and other primary healthcare settings without requiring separate Z-Score calculation tools.

II. Research Method

This study used a comparative quantitative design based on controlled machine learning experiments. The data source consisted of four CSV files from Jeneponto Regency covering 2021, 2022, 2023, and 2024. The combined raw dataset contained 40,072 rows. One blank row was removed, leaving 40,071 labeled toddler records for modeling. The target variable was height-for-age status, coded as stunted and not stunted. All Z-Score columns and derived nutritional indicators were excluded from the predictor set.

The final input features were age in months, sex, weight, and height. Sex was encoded as a binary variable. Age and weight contained a small number of missing values, specifically two missing age values and one missing weight value. These values were handled through median imputation fitted inside the training pipeline. Several obvious weight unit-entry errors were corrected before modeling: newborn values recorded as 2,800-3,100 were converted to 2.8-3.1 kg, and one value of 91 was converted to 9.1 kg. After this correction, weight ranged from 2.0 to 28.5 kg, which is clinically more plausible for the dataset.

The dataset was divided into training and testing subsets using a stratified 80:20 split with random_state=42. Stratification preserved the stunted and not-stunted class proportions in both subsets. SMOTE was applied only to the training data after the split and inside each cross-validation fold[12]. The testing data were never oversampled. This design prevents oversampling leakage, where synthetic samples generated from the full dataset could indirectly contaminate the test set.

The class distribution consisted of 23,867 not-stunted records (59.56%) and 16,204 stunted records (40.44%). A temporal imbalance was observed: the 2021, 2022, and 2023 files contained only stunted cases, while the not-stunted class appeared in the 2024 file. This does not invalidate the random stratified experiment, but it limits temporal and geographic generalization. The limitation is explicitly discussed because a model intended for wider public-health deployment should be externally validated using data from other years and regions.

This research was executed through systematic and controlled phases to guarantee that the resultant categorization model is valid, quantifiable, and devoid of algorithmic bias. The complete stages of the Machine Learning experimental approach in this research are illustrated in Figure 1 below.

Figure 1 shows that the model training stage evaluates four algorithms rather than a single classifier. The workflow also includes leakage-aware preprocessing, SMOTE only on training data, stratified 5-fold cross-validation, extended metrics, statistical testing, and explainable AI analysis

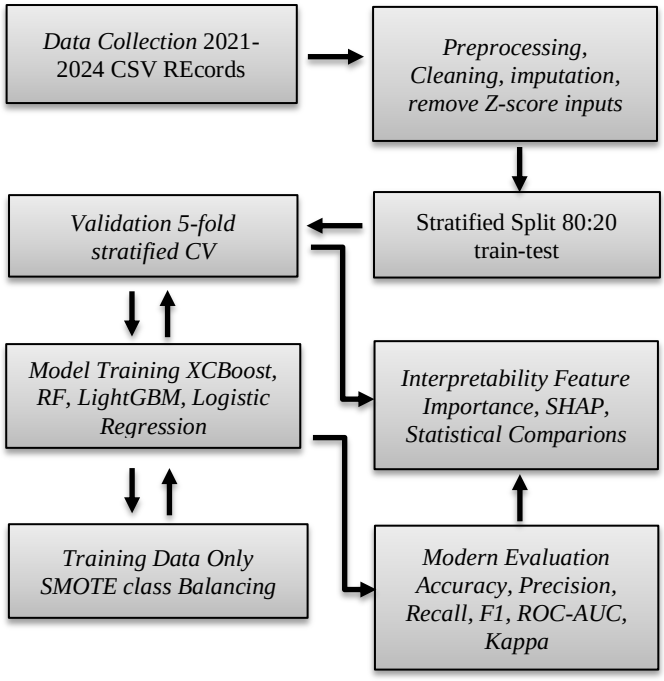

Figure 1. Research flow diagram representing all evaluated algorithms and validation stages.

As shown in Figure 1, every step of the Machine Learning experimental methodology in this study is carried out visually through a methodical process to guarantee that the final classification model is valid, measurable, robust against anomalies in the class distribution, and free from algorithmic bias. Six essential, interrelated phases make up the design of the research flow. Data collection, which starts the first phase, involves extracting the historical secondary anthropometric dataset of infants between the ages of 0 and 59 months. After that, the raw data is sent to the second stage, Data Preprocessing, which comprises label encoding modification for categorical variables, outlier cleaning, missing value imputation, and the rigorous removal of all derived calculation features, including the WHO's built-in Z-Score[13] . In order to avoid data leakage, which frequently causes a false sense of optimism bias in medical categorization predictive modeling, it is essential to remove these generated features[14] . In order to preserve the percentage of class population representation, the third phase, Data Splitting, divides the resulting clean dataset using a stratified splitting strategy with an 80% ratio for training and 20% for testing.

2.1. Experimental Setup and Model Configuration

All experiments were conducted using Python 3.11 with Pandas, NumPy, Scikit-learn, Imbalanced-learn, XGBoost, LightGBM, and SHAP. The experiment used CPU-based computation so that the workflow can be reproduced on ordinary computing devices without GPU acceleration.

The XGBoost model was configured using n_estimators=300, learning_rate=0.05, max_depth=5, subsample=0.8, colsample_bytree=0.8, objective="binary:logistic", eval_metric="logloss", tree_method="hist", and random_state=42. Random Forest used n_estimators=300, criterion="gini", max_depth=None, min_samples_split=2, min_samples_leaf=1, bootstrap=True, and random_state=42. LightGBM used n_estimators=300, learning_rate=0.05, num_leaves=31, max_depth=-1, subsample=0.8, colsample_bytree=0.8, objective="binary", and random_state=42. Logistic Regression used solver="liblinear", penalty="l2", C=1.0, and random_state=42.

The hyperparameters in this revision were fixed, manually specified configurations rather than values obtained through Grid Search, Random Search, or Bayesian Optimization[15]. They were selected as common regularized baseline settings for comparative evaluation. Therefore, the reported performance should be interpreted as fixed-configuration performance, not as the globally optimal performance of each algorithm. A future version can further strengthen the study by applying nested cross-validation with hyperparameter search inside each training fold.

2.2 Validation Strategy and Evaluation Metrics

In addition to the 80:20 hold-out evaluation, stratified 5-fold cross-validation was conducted. In every fold, imputation and SMOTE were fitted only on the training portion, while the validation fold was kept in its original distribution. This strategy provides a more reliable estimate of model robustness than a single split alone. The evaluation metrics include accuracy, precision, recall, F1-score, ROC-AUC, Matthews Correlation Coefficient (MCC), and Cohen's Kappa[16]. Precision, recall, and F1-score were calculated with the stunted class as the positive class. In medical screening, recall is important because false negatives may delay intervention, while MCC and Kappa provide more balanced assessment when classes are imbalanced.

2.3 Statistical Comparison

Statistical comparison was performed using two approaches. First, the Friedman test was applied to the 5-fold F1-scores to evaluate whether the algorithms differed overall. Second, pairwise Wilcoxon signed-rank tests compared XGBoost against the other models using fold-level F1-scores. In addition, an exact paired McNemar-style binomial test was applied to hold-out prediction correctness to compare whether two classifiers made significantly different errors on the same test records.

Feature interpretation was strengthened using both model-based feature importance and SHAP. SHAP was applied to the XGBoost model because XGBoost was the original focus of the proposed framework and provides stable tree-based

explanations[17]. The SHAP summary plot was used to identify the magnitude of feature contribution, while the dependence plot was used to examine how height and age jointly influence the predicted probability of stunting.

III. Results and Discusion

3.1 Dataset Characteristics and Exploratory Data Analysis

The cleaned dataset consisted of 40,071 labeled records. The not-stunted class accounted for 23,867 records (59.56%), while the stunted class accounted for 16,204 records (40.44%). This distribution shows moderate imbalance, justifying the use of SMOTE on training data.

The temporal class distribution was uneven. All records in 2021, 2022, and 2023 were labeled stunted, while the 2024 data contained both stunted and not-stunted children. This pattern suggests that the dataset may reflect different collection procedures or reporting scope across years. Therefore, although the random stratified validation is useful for algorithm comparison, external validation remains necessary before claiming broader deployment readiness.

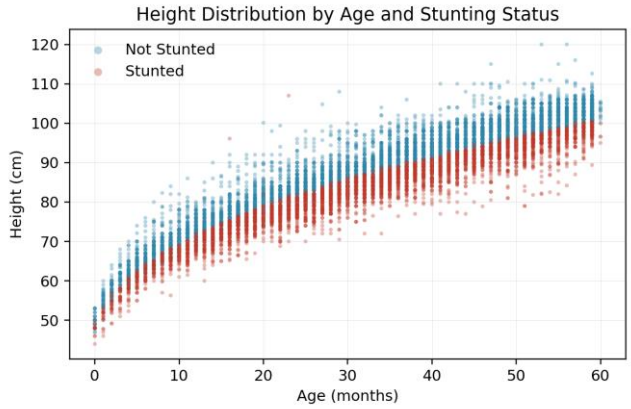

Figure 2: Height distribution by age according to stunting status.

Figure 2 shows the biological separation between the two classes. Stunted children tend to occupy the lower height range at comparable ages, which is consistent with the clinical definition of stunting as impaired linear growth relative to age. However, overlap still appears in some age-height regions, indicating that the classification boundary is not perfectly linear and may benefit from tree-based learning.

Figure 3 below indicates strong correlations among age, weight, and height. Height was strongly correlated with age (r=0.93) and weight (r=0.94), while age and weight were also strongly correlated (r=0.88). This multicollinearity can reduce the stability of simple linear models, but tree-based ensembles can model non-linear interactions without requiring dimensionality reduction.

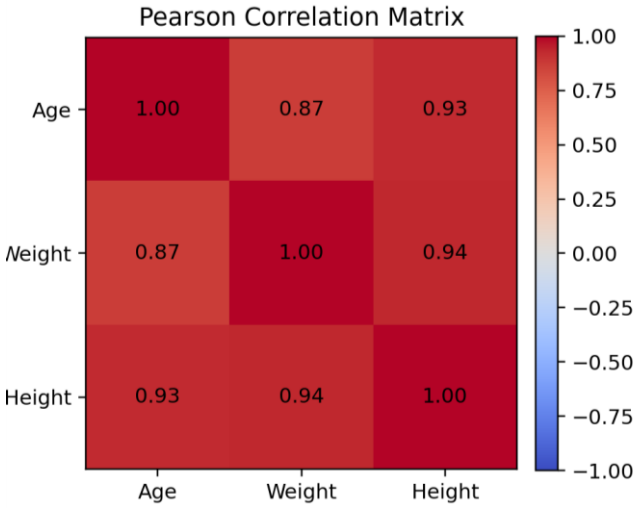

Figure 3. Correlation matrix of basic anthropometric variables

3.2. Hold-out Model Performance

Table 1 presents the expanded hold-out performance metrics requested in the review. The results are reported using the stunted class as the positive class for precision, recall, and F1-score

Table 1. Extended comparative hold-out performance of stunting prediction models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)	MCC	Kappa
Random Forest	95.81	94.00	95.74	94.86	99.09	0.913	0.913
LightGBM	95.72	93.72	95.83	94.77	99.28	0.912	0.911
XGBoost	94.98	92.39	95.46	93.90	99.12	0.897	0.896
Logistic Regression	84.22	77.99	84.94	81.32	91.53	0.679	0.677

The updated results differ from the earlier manuscript claim. Under the valid Z-Score-free pipeline, Random Forest achieved the highest hold-out accuracy (95.81%), F1-score (94.86%), MCC (0.913), and Cohen's Kappa (0.913). LightGBM achieved the highest recall (95.83%) and ROC-AUC (99.28%). XGBoost remained highly competitive, with 95.00% accuracy, 95.46% recall, and 99.12% ROC-AUC, but it did not outperform Random Forest and LightGBM on the main threshold-based metrics.

This finding is scientifically important. It indicates that the previous statement that XGBoost was the best model is not supported after recalculation using the attached source data and a leakage-aware validation pipeline. The manuscript should

therefore frame XGBoost as a strong gradient-boosting benchmark rather than as the single superior classifier. For deployment, Random Forest or LightGBM should also be considered and externally validated.

The lower Logistic Regression performance is biologically plausible. Stunting classification depends on the interaction between age and height, and the same height can imply different nutritional status at different ages[18]. Logistic Regression relies on a more linear decision boundary, so it is less able to capture the age-height-weight interaction that defines child growth patterns. Tree-based models performed better because they can split the feature space into non-linear age- and height-specific regions.

3.3 Stratified 5-fold Cross-validation

Table 2 reports mean and standard deviation across stratified 5-fold cross-validation. SMOTE and imputation were fitted only inside each training fold.

Table 2. Stratified 5-fold cross-validation results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	ROC-AUC (%)	MCC	Kappa
Random Forest	95.81 ± 0.28	94.51 ± 0.38	95.17 ± 0.34	94.84 ± 0.34	99.07 ± 0.12	0.913 ± 0.006	0.913 ± 0.006
LightGBM	95.54 ± 0.23	93.58 ± 0.29	95.51 ± 0.40	94.54 ± 0.28	99.22 ± 0.07	0.908 ± 0.005	0.908 ± 0.005
XGBoost	94.96 ± 0.28	92.46 ± 0.34	95.30 ± 0.54	93.86 ± 0.35	99.07 ± 0.06	0.896 ± 0.006	0.896 ± 0.006
Logistic Regression	84.18 ± 0.58	77.93 ± 0.68	84.94 ± 0.75	81.29 ± 0.68	91.57 ± 0.50	0.678 ± 0.012	0.676 ± 0.012

The cross-validation results are consistent with the hold-out results. Random Forest produced the highest mean F1-score (94.84% ± 0.34), accuracy (95.81% ± 0.28), MCC (0.913 ± 0.006), and Kappa (0.913 ± 0.006). LightGBM produced the highest mean recall (95.51% ± 0.40) and ROC-AUC (99.22% ± 0.07). XGBoost remained close, but its mean F1-score (93.86% ± 0.35) was lower than Random Forest and LightGBM.

3.4 Feature Importance and SHAP-based Interpretation

The XGBoost gain-based importance shows that height (42.34%) and age (34.84%) were the largest contributors, followed by weight (12.05%) and sex (10.77%). SHAP produced an even clearer interpretation: height accounted for 50.54% of the mean absolute SHAP contribution and age accounted for 37.04%. This pattern is clinically coherent because stunting is fundamentally a deficit in linear growth relative to age

Table 4. XGBoost feature importance and SHAP contribution

Feature	XGB Gain Importance (%)	Mean |SHAP| Contribution (%)
Height	42.34	50.54
Age	34.84	37.04
Weight	12.05	3.83
Sex	10.77	8.60

Figure 4 indicates that height and age dominate individual predictions. Lower height values generally push predictions toward the stunted class, while the effect of age depends on the corresponding height pattern. This supports the claim that the model is learning a biologically meaningful approximation from raw anthropometric measurements rather than directly using Z-Score-derived predictors.

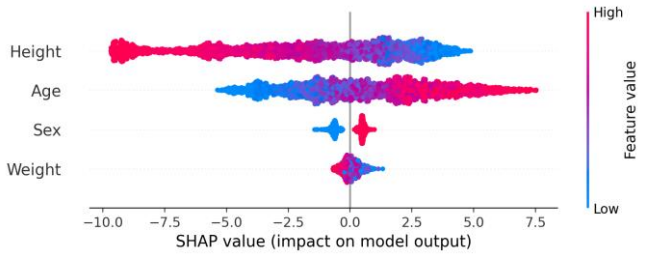

Figure 5. SHAP dependence plot showing the effect of height with age interaction.

Figure 5 shows that the contribution of height is not isolated from age. Similar height values can carry different risk implications depending on the child's age, which explains why non-linear tree-based models outperform Logistic Regression. This interaction reflects the biological logic of child growth assessment, where height must always be interpreted relative to age and sex

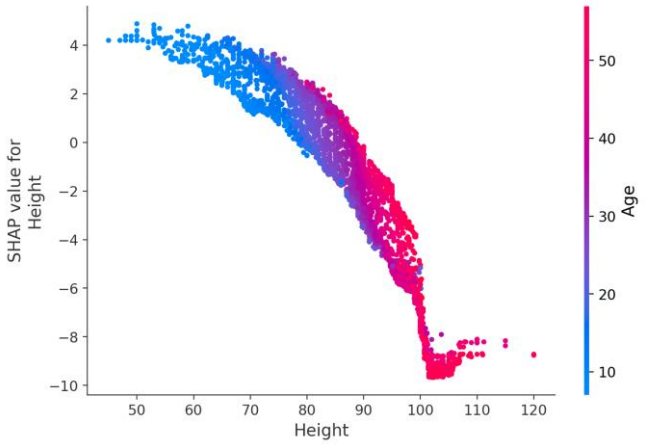

Figure 5. SHAP dependence plot showing the effect of height with age interaction

3.6 Comparison with Previous Studies and Novelty Positioning

The results are consistent with previous studies showing that ensemble learning is useful for stunting prediction. However, the proposed study differs methodologically by explicitly isolating the feature set from deterministic diagnostic variables. Some prior studies and public datasets include Z-Score-derived columns or nutritional-status labels such as height-for-age status, weight-for-age status, and weight-for-height status. These variables are useful for clinical assessment, but they should not be used as predictors when the target is stunting status because they encode information that is directly related to the label.

The novelty of this work is therefore not simply the use of XGBoost or Random Forest. Those algorithms have already been used in previous stunting studies. The novelty lies in the leakage-aware modeling framework: all WHO Z-Score values, height-for-age category, weight-for-age category, weight-for-height category, and other derived nutritional indicators are removed from the input features before model training. This makes the experiment a stricter test of whether machine learning can classify stunting from raw Posyandu-style measurements alone.

This contribution is scientifically important because a model trained with derived diagnostic variables may behave as a formula translator. In contrast, the proposed framework provides a more valid prediction model for practical use: cadres only need to enter age, sex, weight, and height. The model can then provide a rapid screening output without requiring a separate Z-Score calculator as an input dependency.

3.7 Practical Implementation in Posyandu

The proposed framework can be implemented as a lightweight decision-support module in a mobile or web-based Posyandu application. The practical workflow is simple: (1) cadres enter age in months, sex, weight, and height; (2) the system checks input completeness and plausible ranges; (3) the trained model produces a stunting-risk classification and probability score; (4) the application displays a recommendation such as "requires further assessment" or "routine monitoring"; and (5) cases predicted as stunted are confirmed by health workers using the official nutritional assessment procedure.

The computational requirement is low because only four input variables are used. Inference can run on a standard CPU-based server, ordinary laptop, or lightweight web backend. For a single child record, prediction time is expected to be near real-time because Random Forest, LightGBM, and XGBoost require only tabular input and no image, sensor, or deep-learning computation. Batch prediction for Posyandu sessions can also be performed by uploading spreadsheet data.

The system should not replace clinical judgment. It should be positioned as an early screening aid that helps prioritize children for follow-up. This distinction is important ethically and clinically because official stunting diagnosis still depends on standardized measurement procedures, health-worker interpretation, and appropriate nutritional intervention planning[19].

3.8 Limitations and Future Work

First, the dataset comes from a single regency, Jeneponto. Therefore, geographic generalization is limited. Children in other districts may differ in growth patterns, measurement practices, socioeconomic context, sanitation, dietary intake, and health-service access. External validation using datasets from other regions is necessary before deployment at provincial or national scale.

Second, the temporal distribution of the dataset is uneven because the 2021-2023 records contain only stunted cases, while the 2024 data contain both classes. This limits the ability to perform a true temporal validation strategy. Future research should collect balanced yearly datasets and evaluate prospective validation where models trained on earlier years are tested on later years.

Third, this study intentionally used only basic anthropometric measurements to support lightweight deployment. This improves feasibility but limits causal interpretation. Future work can develop two model versions: a lightweight screening model using only age, sex, weight, and height, and an expanded risk-explanation model including dietary intake, birth history, infection history, sanitation, parental education, and socioeconomic variables.

IV. Conclusion

1. The attached source data produce 40,071 valid labeled records after removing one blank row. The dataset contains 23,867 not-stunted records and 16,204 stunted records, with moderate class imbalance and a notable temporal imbalance across years.
2. A leakage-aware machine learning framework was successfully implemented by excluding all WHO Z-Score columns, nutritional-status labels, and derived anthropometric indicators from the input features. The final predictors were only age, sex, weight, and height.
3. After recalculation using the source data, Random Forest achieved the best hold-out accuracy, F1-score, MCC, and Cohen's Kappa, while LightGBM achieved the best ROC-AUC and recall. XGBoost remained highly competitive but did not outperform all other classifiers. Therefore, the manuscript should avoid claiming that XGBoost is conclusively superior on this dataset.
4. Stratified 5-fold cross-validation strengthened the reliability of the findings and showed a consistent performance

pattern. Statistical testing also supports a more cautious interpretation of model differences, especially between the tree-based ensemble models.

5. Feature importance and SHAP analysis show that height and age are the dominant predictors, which is consistent with the biological basis of stunting as impaired linear growth relative to age.
6. The proposed framework has practical potential for Posyandu implementation because it requires only routine field measurements and can run in a lightweight mobile or web-based decision-support system. However, external validation using datasets from other regions is required before broad deployment.

V. References

[1] A. Soliman *et al.*, "Early and long-term consequences of nutritional stunting: From childhood to adulthood," *Acta Biomedica*, vol. 92, no. 1, Mar. 2021, doi: 10.23750/abm.v92i1.11346.

[2] M. Asif *et al.*, "Establishing Height-for-Age Z-Score Growth Reference Curves and Stunting Prevalence in Children and Adolescents in Pakistan," *Int. J. Environ. Res. Public Health*, vol. 19, no. 19, Oct. 2022, doi: 10.3390/ijerph191912630.

[3] I. Y. Tian *et al.*, "3D convolutional deep learning for nonlinear estimation of body composition from whole body morphology," *NPJ Digit. Med.*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s41746-025-01469-6.

[4] S. Mehta *et al.*, "Advances in artificial intelligence and precision nutrition approaches to improve maternal and child health in low resource settings," *Nature Communications* , vol. 16, no. 1, Dec. 2025, doi: 10.1038/s41467-025-62985-3.

[5] G. Carloni, A. Berti, and S. Colantonio, "The Role of Causality in Explainable Artificial Intelligence," Jun. 01, 2025, *John Wiley and Sons Inc*. doi: 10.1002/widm.70015.

[6] N. Hasdyna, R. K. Dinata, Rahmi, and T. I. Fajri, "Hybrid Machine Learning for Stunting Prevalence: A Novel Comprehensive Approach to Its Classification, Prediction, and Clustering Optimization in Aceh, Indonesia," *Informatics*, vol. 11, no. 4, Dec. 2024, doi: 10.3390/informatics11040089.

[7] M. Park, Y. H. Kim, and J. S. Lee, "Machine Learning-Based Model for Grip Strength Prediction in Healthy Adults: A Nationwide Dataset-Based Study," *J. Clin. Med.*, vol. 14, no. 5, Mar. 2025, doi: 10.3390/jcm14051542.

[8] I. Adewumi, O. F. Afe, A. Ayoade, and U. C. Elugbindin, "Machine Learning-Driven Early Detection of Childhood Malnutrition in Low-Resource Settings: Integrating Anthropometric Imaging and Socioeconomic Data," Sep. 03, 2025. doi: 10.20944/preprints202509.0292.v1.

[9] Z. A. Mohammad, "How Ensemble Learning Balances Accuracy and Overfitting: A Bias-Variance Perspective on Tabular Data," Dec. 2025, [Online]. Available: http://arxiv.org/abs/2512.05469

[10] P. K. Arya, K. Sur, T. Kundu, S. Dhote, and S. K. Singh, "Unveiling predictive factors for household-level stunting in India: A machine learning approach using NFHS-5 and satellite-driven data," *Nutrition*, vol. 132, Apr. 2025, doi: 10.1016/j.nut.2024.112674.

[11] S. S. Chai, K. L. Goh, W. L. Cheah, Y. H. R. Chang, and G. W. Ng, "Hypertension Prediction in Adolescents Using Anthropometric Measurements: Do Machine Learning Models Perform Equally Well?," *Applied Sciences (Switzerland)*, vol. 12, no. 3, Feb. 2022, doi: 10.3390/app12031600.

[12] B. Or, "Improving Requirements Classification with SMOTE-Tomek Preprocessing," May 2026, [Online]. Available: http://arxiv.org/abs/2501.06491

[13] P. Koukaras and C. Tjortjis, "Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices," Oct. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ai6100257.

[14] S. Editor Wolfgang Walz, "Machine Learning for Brain Disorders." [Online]. Available: http://www.springer.com/series/7657

[15] A. M. Mohammed, E. Onieva, M. Woźniak, A. Mohammed, and M. Woźniak, "Employing Bayesian Optimization for Hyperparameter Tuning in Convolutional Neural Networks," 2026. [Online]. Available: https://github.com/AmgadMonir/BO-Based-CNN.

[16] M. Imani, A. Beikmohammadi, and H. R. Arabnia, "Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Upsampling Under Varying Imbalance Levels," Jan. 30, 2025. doi: 10.20944/preprints202501.2274.v1.

[17] Y. Meng, N. Yang, Z. Qian, and G. Zhang, "What makes an online review more helpful: An interpretation framework using xgboost and shap values," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 3, pp. 466–490, 2021, doi: 10.3390/jtaer16030029.

[18] H. D. S. Ferreira, "Anthropometric assessment of children's nutritional status: A new approach based on an adaptation of Waterlow's classification," *BMC Pediatr.*, vol. 20, no. 1, Feb. 2020, doi: 10.1186/s12887-020-1940-6.

[19] D. Wanda *et al.*, "Exploring practical issues in children's anthropometric measurements: A qualitative descriptive study involving Indonesian health professionals and community health workers," *Belitung Nurs. J.*, vol. 11, no. 5, pp. 538–546, Sep. 2025, doi: 10.33546/bnj.3987.