Comparative Evaluation of Ensemble Machine Learning Models for Child Stunting Prediction Using Routine Anthropometric Data in Indonesia

Mikha Dayan Sinaga[1*], Ratna Sri Hayati[2], Novriza Rahayu[3]

[1,2,3]*Universitas Satya Terra Bhinneka, Indonesia*

*corr-author: mikhasinaga@satyaterrabhinneka.ac.id

Abstract - Child stunting remains a major public health challenge in Indonesia and continues to hinder progress toward the Sustainable Development Goals (SDGs), particularly in child health and nutrition. Early identification of at-risk children is therefore essential to support timely interventions. While previous machine learning studies on stunting prediction commonly incorporate socioeconomic, environmental, and behavioral variables, comparative evaluations based exclusively on routinely collected anthropometric indicators remain limited, particularly within Indonesian primary healthcare settings. This study evaluates the predictive performance of multiple machine learning models for stunting classification using only anthropometric and early-life growth indicators. A dataset consisting of 1,000 child records—including age, birth weight, birth length, current weight, current length, and breastfeeding status—was analyzed using Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Gradient Boosting algorithms. The dataset was partitioned using an 80:20 stratified train–test split, while five-fold cross-validation was applied during model development to improve robustness and reproducibility. Experimental results demonstrate that ensemble-based methods outperform single classifiers, with Gradient Boosting achieving the highest predictive performance (accuracy = 0.90, F1-score = 0.90, AUC = 0.93). Feature importance analysis reveals that birth length, birth weight, current weight, and age are among the most influential predictors of stunting risk. These findings suggest that machine learning models built solely on routinely collected anthropometric indicators can provide a practical, scalable, and data-driven approach for early stunting detection in Indonesian primary healthcare systems.

Keywords: Child stunting, machine learning, feature importance, public health prediction, Indonesia.

I. INTRODUCTION

Child stunting remains a critical public health issue in Indonesia and continues to pose a substantial challenge to national health development efforts [1]. Stunting is characterized by chronic impairment in linear growth and is widely recognized as an indicator of long-term nutritional deprivation and repeated exposure to adverse health conditions. The condition is typically assessed using height-for-age indices in accordance with the World Health Organization (WHO) Child Growth Standards [2-4] . Children are classified as stunted when their height-for-age Z-score falls below -2 standard deviations from the reference population [5, 6]. The high prevalence of stunting in Indonesia reflects persistent inequalities in maternal and child nutrition and underscores the need for effective preventive strategies.

The implications of stunting extend beyond physical growth retardation and have been extensively documented in the literature. Stunted children are at increased risk of morbidity and mortality during early childhood and often experience delayed cognitive development, impaired educational attainment, and reduced economic productivity in adulthood [7]. At the population level, stunting contributes to intergenerational cycles of poverty and undermines human capital formation, thereby hindering broader socioeconomic development and progress toward the Sustainable Development Goals (SDGs), particularly those related to child health, nutrition, and well-being.

Early identification of children at risk of stunting is therefore essential to facilitate timely and targeted nutritional interventions during critical growth periods. In many primary health care settings, especially in low- and middle-income countries, the availability of comprehensive socioeconomic, dietary, or environmental data is often limited due to resource constraints. In contrast, anthropometric measurements are routinely collected as part of maternal and child health surveillance programs, such as growth monitoring and immunization services. As a result, anthropometric indicators represent a practical, cost-effective, and scalable data source for early stunting risk detection [8, 9].

Previous studies on stunting prediction have predominantly employed conventional statistical techniques, including linear regression and logistic regression models, using anthropometric and demographic variables [10-12]. While these approaches have contributed valuable insights into risk factors associated with stunting, they are generally constrained by assumptions of linearity and independence among predictors. Such limitations may hinder their ability to capture the complex, nonlinear interactions between early-life growth indicators, current nutritional status, and feeding practices that collectively influence child growth outcomes [13, 14]. Given the multifactorial and dynamic nature of stunting, more flexible analytical frameworks are therefore required.

In recent years, machine learning (ML) methods have gained increasing attention in public health research due to their capacity to model nonlinear relationships, handle multicollinearity, and uncover latent patterns within complex datasets [15-17] . ML algorithms have been successfully applied to various health-related classification and prediction tasks, including disease diagnosis, risk stratification, and nutritional status assessment. Several recent studies have explored the application of machine learning techniques for predicting child malnutrition and stunting in different countries using diverse sets of demographic, socioeconomic, and health-related variables. While these approaches have demonstrated promising predictive performance, many of them rely on extensive datasets that may not always be available in routine public health monitoring systems. [3, 18, 19].

Furthermore, most existing studies focus on specific machine learning algorithms or utilize complex feature sets that combine socioeconomic, environmental, and behavioral variables. Although such variables can enhance predictive performance, they are often difficult to obtain consistently in primary healthcare environments, particularly in resource-constrained settings. In contrast, anthropometric measurements such as birth weight, birth length, body weight, and body length are systematically collected during maternal and child health services and therefore represent one of the most reliable sources of longitudinal child growth information. However, Table I shows the systematic comparative evaluations of multiple machine learning algorithms that rely exclusively on these routinely collected anthropometric indicators remain relatively limited, particularly within the Indonesian healthcare context.

Table I summarizes recent studies on machine learning-based stunting prediction. Existing research demonstrates that machine learning models can achieve strong predictive performance, particularly when leveraging large-scale datasets with diverse socioeconomic, demographic, and health-related features. However, such data are often not readily available in routine primary healthcare settings, especially in resource-constrained environments. A limited number of studies have explored the use of anthropometric indicators alone, and even fewer have conducted systematic comparisons of multiple machine learning algorithms using exclusively these routinely collected variables within the Indonesian context. Therefore, this study addresses this gap by evaluating and comparing multiple machine learning models based solely on anthropometric data, providing a more practical and scalable approach for early stunting detection in primary healthcare systems.

TABLE I
STATE-OF-THE-ART COMPARISON OF MACHINE LEARNING APPROACHES FOR STUNTING PREDICTION

Study / Context	Country	Dataset	Features	Methods	Key Findings
ML for stunting detection (Centing App) [7]	Indonesia	Not specified	Anthropometric + app data	Multiple ML	Demonstrates feasibility of ML in mobile-based stunting detection
LightGBM stunting classification [8]	Indonesia	Regional dataset	Anthropometric	LightGBM	High classification accuracy using boosting methods
ML for stunting prediction [9]	Egypt	Large survey data	Socioeconomic + health	Supervised ML	Strong predictive performance using complex features
Hybrid ML for stunting [12]	Indonesia	Regional dataset	Mixed variables	Hybrid ML	Improved accuracy through hybrid modeling
Comparative ML using routine healthcare data (This Study)	**Indonesia**	**1,000 records**	**Anthropometric only**	**LR, DT, RF, SVM, GB**	**Systematic comparison; practical for primary healthcare**

Accordingly, this study aims to evaluate the predictive performance of multiple machine learning algorithms for identifying child stunting in Indonesia based exclusively on anthropometric and early-life growth indicators, including age, birth length, birth weight, current height, current weight, and breastfeeding status. By comparing the performance of both single and ensemble-based machine learning models, this study seeks to identify the most effective approaches for early stunting detection using routinely available healthcare data. The findings are expected to contribute to the development of practical, data-driven tools that can be integrated into routine child health monitoring systems, thereby supporting evidence-based decision-making and strengthening stunting prevention efforts in Indonesia.

II. METHOD

To enhance methodological clarity and provide a comprehensive overview of the research process, a machine learning workflow diagram is presented in Fig. I. The diagram illustrates the sequential stages of the study, beginning with raw anthropometric data collection, followed by data preprocessing procedures, including transformation and encoding. It then outlines the stratified train–test split applied to the dataset, the model training phase involving multiple machine learning algorithms, and the subsequent performance evaluation using standard classification metrics. Finally, the workflow highlights the feature importance analysis conducted to identify key predictors of stunting. This visual representation is intended to improve transparency, facilitate reproducibility, and provide readers with a clear understanding of the overall analytical pipeline employed in this study.

A. Data Source and Variables

This study utilized anthropometric data obtained from routine child growth monitoring conducted in primary healthcare services in Indonesia. These records originate from standard maternal and child health surveillance activities in which children's growth indicators are periodically measured and documented by healthcare workers. Because the dataset was derived from routine health monitoring practices, it reflects real-world measurement conditions commonly encountered in community-based healthcare environments.

The dataset consists of 1,000 individual child observations, each containing a set of anthropometric and early-life indicators associated with child growth and nutritional status. The variables include child age, birth weight, birth length, current body weight, current body length (or height), breastfeeding status, and stunting status as the target variable. These variables represent both baseline growth conditions at birth and current physical development, enabling the model to capture growth patterns associated with stunting risk.

Anthropometric indicators were intentionally selected because they represent routinely available measurements in primary healthcare systems, making them suitable for scalable early screening tools. Unlike studies that require complex socioeconomic or dietary information, the use of anthropometric indicators enables the development of practical machine learning models that can be easily integrated into routine health monitoring systems.

To illustrate the structure of the dataset, a representative subset of the raw records is presented in Table II, demonstrating the format and measurement characteristics of the original data prior to preprocessing.

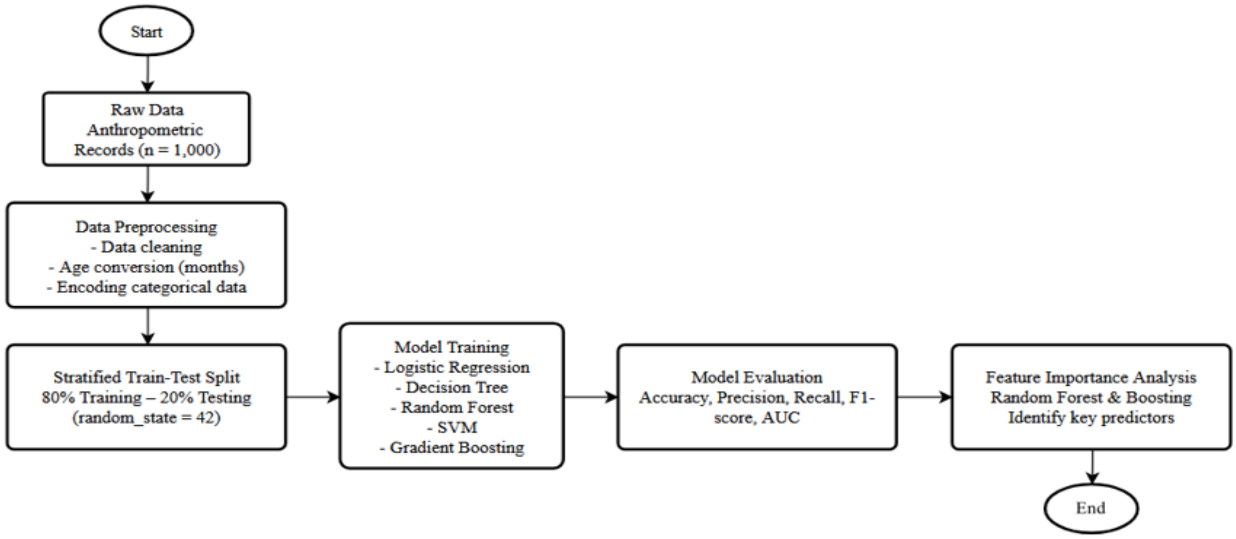

Fig. 1 Machine learning workflow for child stunting prediction

TABLE II
SAMPLE OF RAW ANTHROPOMETRIC DATA

Age	Birthday	Birth Weight (Kg)	Birth Length (cm)	Current Weight (Kg)	Current Length (cm)	Breastfeeding Status	Stunting Status
1 Y - 7 M - 16 D	25/06/2023	3	49	10	77	Yes	Yes
1 Y - 4 M - 23 D	18/09/2023	2.7	49	8.5	75	No	Yes
2 Y - 6 M - 7 D	05/08/2022	3.3	50	10.8	85	Yes	Yes
1 Y - 0 M - 22 D	29/01/2024	2.8	50	7.3	69	Yes	Yes
2 Y - 5 M- 9 D	06/09/2022	3	49	10.1	83.1	Yes	Yes
2 Y - 1 M - 19 D	27/12/2022	2.9	49	10.7	80.5	No	Yes
1 Y - 11 M - 7 D	10/03/2023	3	50	10.6	81	Yes	No
2 Y - 11 M - 5 D	12/03/2022	3.1	50	11	85.5	No	Yes
0 Y - 6 M - 14 D	01/08/2024	2.7	50	6.5	63	Yes	Yes
1 Y - 5 M - 12 D	08/09/2023	3	50	7.9	73	No	No

B. Data Preprocessing

The raw anthropometric data collected from routine health monitoring records required preprocessing before being used in machine learning models. Several inconsistencies in the dataset-such as non-numeric representations and mixed formatting-necessitated systematic transformation to ensure computational compatibility and analytical consistency [15, 20, 21].

First, the child age variable, originally recorded using a composite format of years, months, and days, was converted into a continuous numerical variable representing age in months. This transformation allowed age to be represented consistently across all observations.

Second, numerical values that used comma-based decimal notation were standardized to decimal-point format to ensure compatibility with numerical processing during model training.

Third, categorical variables, including breastfeeding status and stunting status, were encoded into numerical form using binary encoding, where “Yes” was represented as 1 and “No” as 0. This encoding enabled these variables to be processed effectively by machine learning algorithms.

Additionally, the date of birth variable was removed after age conversion because it no longer provided independent informational value and could introduce redundancy in the feature set.

Following preprocessing, the dataset consisted of six predictive variables and one target variable suitable for machine learning analysis. Table III illustrates the transformation of the dataset before and after preprocessing.

C. Experimental Design

To promote methodological transparency and ensure reproducibility, a well-defined and systematic experimental framework was established for model development and evaluation. The dataset was partitioned into training and testing subsets using an 80:20 split ratio, whereby 80% of the data were utilized for model training and the remaining 20% were reserved for independent performance evaluation. To maintain the original distribution of the target variable, stratified sampling was applied during the splitting process, ensuring that the proportion of stunting and non-stunting cases remained consistent across both subsets and reducing potential sampling bias.

All experiments were conducted using a fixed random seed (random_state = 42) to ensure that the results are fully reproducible and consistent across repeated runs. In addition to the hold-out validation approach, five-fold cross-validation was implemented during the training phase to provide more stable and reliable estimates of model performance. This procedure allows the model to be trained and evaluated across multiple data partitions, thereby improving generalization capability and minimizing the risk of overfitting. Collectively, these methodological choices strengthen the robustness, reliability, and credibility of the experimental results presented in this study.

The overall experimental pipeline consisted of the following stages:

1. Raw data acquisition
2. Data preprocessing and feature transformation
3. Stratified train–test data splitting
4. Model training with cross-validation
5. Hyperparameter tuning
6. Model evaluation on the hold-out test set
7. Feature importance analysis

This pipeline ensures a transparent and systematic framework for evaluating the predictive capability of machine learning models in stunting detection.

D. Machine Learning Models

Five machine learning models were implemented to predict stunting status:

1) Logistic Regression: Logistic Regression is a widely used binary classification method that estimates the probability of class membership based on input variables. Due to its interpretability and robustness, it is commonly employed in health-related studies and serves as a baseline model for evaluating the relationship between anthropometric indicators and stunting outcomes [22].

2) Decision Tree: Decision Tree is a supervised learning algorithm that recursively partitions the feature space into decision rules based on criteria such as information gain or Gini impurity. Its interpretable structure facilitates understanding of classification decisions; however, single-tree models are prone to overfitting and high variance [23].

3) Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees constructed using bootstrap sampling and random feature selection. By aggregating predictions through majority voting, it improves predictive accuracy, reduces overfitting, and provides feature importance estimates that are useful for identifying key stunting-related factors [24].

4) Support Vector Machine: Support Vector Machine (SVM) is a supervised learning algorithm that identifies an optimal decision boundary by maximizing the margin between classes. Through the use of kernel functions, SVM can model nonlinear relationships and has demonstrated strong generalization performance in biomedical and public health classification tasks [25], [26].

5) Gradient Boosting: Gradient Boosting is an ensemble technique that sequentially builds models to minimize prediction errors from previous iterations. Its ability to capture complex nonlinear patterns has made it highly effective for structured health datasets. In this study, Gradient Boosting achieved the highest predictive performance among all evaluated models [27].

These models were selected to represent both linear and non-linear classification approaches commonly applied in health prediction studies [12], [28-31].

TABLE III
SAMPLE DATA BEFORE AND AFTER PREPROCESSING

(a) Raw Data (Before Preprocessing)

Age (Text Format)	Birth Weight (kg)	Birth Length (cm)	Current Weight (kg)	Current Length (cm)	Breastfeeding	Stunting
1 Y - 7 M - 16 D	3.0	49	10.0	77.0	Yes	Yes
1 Y - 4 M - 23 D	2.7	49	8.5	75.0	No	Yes
2 Y - 6 M - 7 D	3.3	50	10.8	85.0	Yes	Yes
1 Y - 0 M - 22 D	2.8	50	7.3	69.0	Yes	Yes
2 Y - 5 M - 9 D	3.0	49	10.1	83.1	Yes	Yes
…	…	…	…	…	…	…
1 Y - 5 M - 12 D	3.0	50	7.9	73	No	No

(b) Processed Data (After Preprocessing)

Age (Months)	Birth Weight (kg)	Birth Length (cm)	Current Weight (kg)	Current Length (cm)	Breastfeeding (0/1)	Stunting (0/1)
19	3.0	49	10.0	77.0	Yes	Yes
17	2.7	49	8.5	75.0	No	Yes
30	3.3	50	10.8	85.0	Yes	Yes
12	2.8	50	7.3	69.0	Yes	Yes
29	3.0	49	10.1	83.1	Yes	Yes
…	…	…	…	…	…	…
17	3.0	50	7.9	73	No	No

E. Hyperparameter Tuning

To optimize model performance, hyperparameter tuning was conducted for each machine learning algorithm using grid search combined with five-fold cross-validation. The tuning process explored a range of parameter values to identify configurations that yielded the best predictive performance.

The main hyperparameters considered in this study include:

1) *Gradient Boosting*
   - Number of estimators: 100, 200, 300
   - Learning rate: 0.01, 0.05, 0.1
   - Maximum tree depth: 3, 5, 7
2) *Random Forest*
   - Number of trees (n_estimators): 100, 200
   - Maximum depth: 5, 10, None
3) *Decision Tree*
   - Maximum depth: 3, 5, 10
   - Minimum samples per split: 2, 5
4) *Support Vector Machine*
   - Kernel type: linear, radial basis function (RBF)
   - Regularization parameter (C): 0.1, 1, 10
5) *Logistic Regression*
   - Regularization strength (C): 0.1, 1, 10
   - Solver: lbfgs

The optimal parameter combinations identified during cross-validation were subsequently used to train the final models evaluated on the hold-out test dataset.

F. Class Distribution and Imbalance Handling

Class imbalance is a common challenge in health-related classification problems and can significantly influence model performance. Therefore, the class distribution of the target variable was examined prior to model training.

In this dataset, the proportion of stunting and non-stunting cases was relatively balanced, reducing the risk of severe class imbalance bias. Nevertheless, stratified sampling was applied during dataset splitting to maintain consistent class proportions in both training and testing subsets.

Additionally, performance evaluation relied on multiple metrics—including precision, recall, F1-score, and AUC-rather than accuracy alone, ensuring that model performance was assessed comprehensively and was not biased toward the majority class.

G. Model Evaluation

Model performance was evaluated using a set of widely recognized classification metrics to provide a comprehensive and balanced assessment of predictive capability. These metrics include accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC). Accuracy reflects the overall proportion of correctly classified instances and provides a general indication of model performance. However, relying solely on accuracy may be insufficient, particularly in cases where class distribution is imbalanced.

To address this limitation, precision and recall were also considered. Precision measures the proportion of correctly predicted positive cases (stunting) among all predicted positives, thereby reflecting the model's ability to minimize false positives. In contrast, recall (sensitivity) evaluates the proportion of actual positive cases that are correctly identified, highlighting the model's effectiveness in reducing false negatives. The F1-score, which represents the harmonic mean of precision and recall, provides a balanced metric that is especially useful when both false positives and false negatives carry important implications.

Additionally, the ROC-AUC metric was employed to assess the model's discriminative ability across various classification thresholds. By summarizing the trade-off between true positive and false positive rates, AUC offers a threshold-independent measure of how well the model distinguishes between stunted and non-stunted children, thereby providing a more robust evaluation of overall classification performance [32, 33].

H. Feature Importance Analysis

To identify the anthropometric indicators most influential in stunting prediction, feature importance analysis was performed using Random Forest and Gradient Boosting models. These ensemble methods can capture complex nonlinear relationships while providing interpretable importance scores. Feature importance was calculated based on each variable's contribution to reducing prediction error during tree construction, enabling a quantitative ranking of predictors. The resulting rankings highlight key growth-related factors, including birth conditions and current nutritional status, that contribute most strongly to stunting risk. These findings provide practical insights for prioritizing anthropometric indicators in early screening and monitoring programs, particularly in resource-constrained healthcare settings, and support the development of more targeted and effective stunting prevention strategies.

I. Study Limitations

Several limitations of this study should be acknowledged. First, although the dataset size of 1,000 observations is adequate for classical machine learning analysis, a larger sample would improve model robustness and generalizability. Second, the analysis was restricted to anthropometric and early-life growth indicators and did not include other important determinants of stunting, such as socioeconomic conditions, parental education, dietary intake, sanitation, and environmental factors. Third, external validation using independent datasets from different regions was not conducted, limiting the ability to assess model performance across diverse populations. Fourth, while five-fold cross-validation was implemented to improve reliability, statistical significance testing, confidence interval estimation, and repeated validation procedures were beyond the scope of this study. Consequently, the reported performance differences among models should be interpreted as indicative rather than conclusive. Future research should address these limitations by incorporating larger multicenter datasets, broader feature sets, and more rigorous statistical validation procedures.

III. RESULT AND DISCUSSION

A. Model Performance Evaluation

The predictive performance of the proposed machine learning models was evaluated using a stratified 80:20 hold-out testing strategy, with five-fold cross-validation applied during the training phase to improve robustness and reduce variance in performance estimation. While the dataset size (n = 1,000) is considered moderate and suitable for classical machine learning methods, careful interpretation of model performance remains essential.

A total of five classification algorithms-Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Gradient Boosting-were evaluated using multiple performance metrics, including accuracy, precision, recall, F1-score, and AUC. The comparative results obtained from the hold-out test dataset are presented in Table IV.

The results indicate that Logistic Regression achieved moderate performance, reflecting the limitations of linear models in capturing complex nonlinear relationships among anthropometric variables. The Decision Tree model showed slight improvement; however, its performance remained constrained by known issues of instability and overfitting associated with single-tree structures.

Ensemble-based methods, including Random Forest and Gradient Boosting, demonstrated consistently higher performance across all evaluation metrics. Random Forest achieved an accuracy of 0.88 and an AUC of 0.91, indicating strong predictive capability and generalization performance. Similarly, Gradient Boosting produced the highest numerical performance, with an accuracy of 0.90, an F1-score of 0.90, and an AUC of 0.93.

However, these results should be interpreted with caution. The observed performance difference between Random Forest and Gradient Boosting is relatively small (e.g., 0.88 vs. 0.90 in accuracy) and may not necessarily indicate a statistically significant improvement. In this study, formal statistical significance testing—such as McNemar's test or paired t-test—was not conducted, and confidence intervals or standard deviations were not reported. Therefore, it cannot be conclusively stated that Gradient Boosting outperforms Random Forest in a statistically robust manner.

Overall, the findings suggest that ensemble learning methods tend to offer more reliable predictive performance compared with single classifiers; nevertheless, conclusions regarding model superiority should be interpreted conservatively given the absence of formal statistical validation.

B. ROC Curve and AUC Comparison

To further evaluate the discriminative performance of the classification models, Receiver Operating Characteristic (ROC) curves were generated for each algorithm based on the hold-out testing dataset. Fig. 2 presents the comparison of AUC values across all models, while Fig. 3 illustrates the ROC curves.

TABLE IV
PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS FOR STUNTING PREDICTION

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.78	0.76	0.79	0.77	0.80
Decision Tree	0.81	0.80	0.82	0.81	0.83
Random Forest	0.88	0.87	0.89	0.88	0.91
Support Vector Machine	0.84	0.83	0.85	0.84	0.87
Gradient Boosting	0.90	0.89	0.91	0.90	0.93

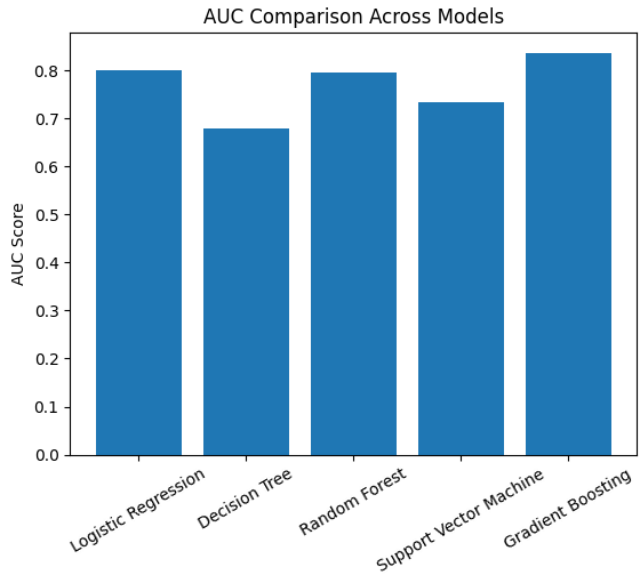

Fig. 2 The comparison of AUC across models

As shown in Fig. 2, ensemble-based methods-particularly Random Forest and Gradient Boosting-achieved the highest AUC values among the evaluated models. These results suggest stronger discriminative capability in distinguishing between stunted and non-stunted children across varying classification thresholds. Logistic Regression and Support Vector Machine (SVM) demonstrated moderate AUC values, while the Decision Tree model exhibited the lowest performance, consistent with its known limitations in terms of generalization and stability.

However, the differences in AUC values between the top-performing models are relatively small. For instance, the gap between Random Forest and Gradient Boosting is limited, and without statistical significance testing or confidence interval estimation, it cannot be conclusively determined whether these differences represent meaningful performance improvements. Therefore, the observed ranking of models should be interpreted as indicative rather than definitive.

The ROC curves presented in Fig. 3 further illustrate these findings. Ensemble-based models tend to produce curves that are closer to the upper-left corner, indicating higher true positive rates at lower false positive rates. This characteristic is particularly important in public health screening applications, where minimizing false negatives is critical to ensure early detection of children at risk of stunting.

Nevertheless, it should be noted that the ROC analysis presented in this study is based on a single hold-out dataset. Although cross-validation was applied during model training to enhance robustness, additional validation using repeated cross-validation or external datasets would be necessary to confirm the stability and generalizability of these results.

C. Feature Importance Analysis

To comprehensively examine the contribution of individual anthropometric variables to stunting prediction, a feature importance analysis was conducted using two ensemble-based machine learning models: Random Forest and Gradient Boosting. These methods provide quantitative importance scores that reflect the extent to which each feature contributes to reducing classification error during model training. In Random Forest, importance is typically derived from the mean decrease in impurity across multiple trees, while Gradient Boosting evaluates feature contributions based on their role in minimizing the loss function through sequential learning. The resulting feature importance values are presented in Table V.

The analysis indicates that current weight, age in months, and current length are among the most influential predictors in the Random Forest model, emphasizing the importance of present nutritional status and cumulative growth patterns. In contrast, the Gradient Boosting model identifies birth length as the most dominant predictor, followed by birth weight and age, highlighting the critical role of early-life growth conditions in shaping long-term developmental outcomes.

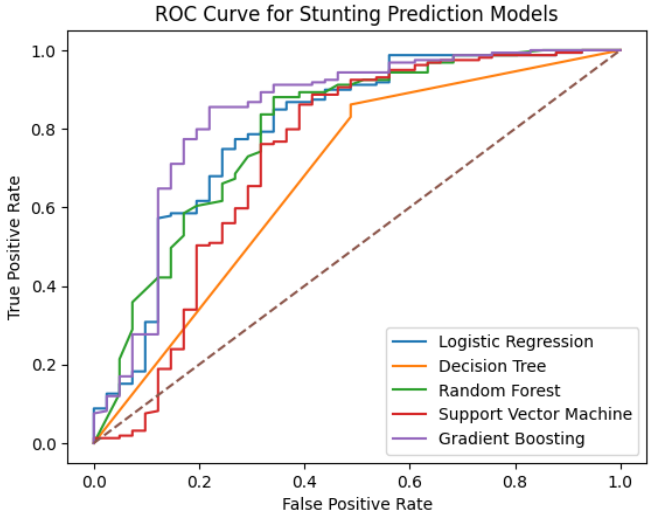

Fig. 3 ROC curve for stunting prediction models

TABLE V
FEATURE IMPORTANCE ANALYSIS USING RANDOM FOREST AND GRADIENT BOOSTING MODELS

No.	Feature	Random Forest Importance	Gradient Boosting Importance
1	Current Weight	0.228	0.187
2	Age (Months)	0.223	0.179
3	Current Length	0.185	0.085
4	Birth Weight	0.173	0.182
5	Birth Length	0.145	0.345
6	Breastfeeding	0.046	0.022

These findings reinforce the understanding of stunting as a multifactorial condition shaped by both prenatal and postnatal factors. The strong influence of birth-related indicators supports the importance of the first 1,000 days of life as a critical period for growth and development. Differences in feature importance rankings between Random Forest and Gradient Boosting further suggest that predictor relevance may vary according to model structure, highlighting the value of employing multiple algorithms for a more comprehensive interpretation.

However, feature importance scores are model-dependent and should not be interpreted as evidence of causal relationships. The exclusion of socioeconomic, environmental, and dietary variables also limits the ability to capture the full range of factors associated with stunting. In addition, the stability of the importance rankings was not formally evaluated through statistical inference or resampling methods. Despite these limitations, the results demonstrate the practical utility of routinely collected anthropometric data for early stunting detection in resource-constrained healthcare settings. Future studies should incorporate broader determinants, validate findings using external datasets, and explore causal inference approaches to improve the robustness and generalizability of machine learning-based stunting prediction models.

D. Comparison with Previous Studies and Implications

The findings of this study are generally consistent with previous research demonstrating the effectiveness of machine learning techniques for stunting prediction. Ensemble-based algorithms, particularly Random Forest and Gradient Boosting, achieved the highest predictive performance among the evaluated models, supporting evidence from recent studies that have identified boosting and ensemble methods as highly effective for health-related classification tasks. For example, [8] reported strong classification performance using the LightGBM algorithm for child stunting detection, while [12] demonstrated improved predictive capability through hybrid machine learning approaches in Indonesia.

A notable distinction of the present study is its exclusive reliance on routinely collected anthropometric indicators. In contrast to many previous studies that incorporated socioeconomic, environmental, dietary, or behavioral variables, the proposed approach achieved competitive predictive performance using only variables that are commonly available in primary healthcare services. This finding highlights the practical feasibility of implementing machine learning-based screening tools in resource-constrained settings where comprehensive data collection may not be possible.

Nevertheless, the generalizability of the results should be interpreted cautiously. The dataset used in this study represents a specific population and healthcare context, and model performance may vary when applied to different geographic regions, demographic groups, or healthcare systems. Furthermore, although five-fold cross-validation was employed during model development, external validation using independent datasets was not performed. Therefore, additional studies involving larger and more diverse populations are required to confirm the robustness and transferability of the proposed models.

Another important consideration concerns the potential risk of overfitting. While ensemble methods generally provide strong predictive performance and improved generalization, their complexity may increase sensitivity to dataset-specific patterns. The use of stratified sampling, cross-validation, and hyperparameter tuning was intended to mitigate this risk; however, future studies should incorporate repeated cross-validation, confidence interval estimation, and statistical significance testing to provide stronger evidence regarding comparative model performance.

Overall, the results suggest that routinely collected anthropometric indicators contain sufficient predictive information to support early stunting detection and may serve as the foundation for scalable decision-support systems in Indonesian primary healthcare services.

IV. CONCLUSION

This study investigated the predictive performance of several machine learning algorithms for identifying child stunting in Indonesia using routinely collected anthropometric indicators. The results indicate that ensemble-based models, particularly Gradient Boosting and Random Forest, achieved higher predictive performance compared with single-classifier approaches, reflecting their ability to capture nonlinear relationships in child growth data. Although Gradient Boosting produced the highest numerical performance across evaluation metrics (accuracy 0.90, F1-score 0.90, AUC 0.93), these differences should be interpreted with caution, as statistical significance testing and confidence interval estimation were not conducted. Feature importance analysis suggests that both early-life indicators (e.g., birth length and birth weight) and current anthropometric measures (e.g., weight, length, and age) contribute substantially to stunting prediction, while breastfeeding status appears to have a more indirect role. Nevertheless, several limitations should be explicitly acknowledged. The dataset size (n = 1,000), while adequate for classical machine learning, remains moderate and may limit generalizability. The analysis is restricted to anthropometric variables and does not incorporate broader socioeconomic, environmental, or dietary determinants of stunting. Furthermore, the absence of external validation using independent datasets limits the ability to assess model performance across different populations. Therefore, the findings should be interpreted as indicative rather than definitive. Future research should incorporate larger and more diverse datasets, include additional determinants, and perform external validation and statistical testing to enhance the robustness, generalizability, and practical applicability of machine learning–based stunting prediction models.

ACKNOWLEDGEMENT

The authors express their sincere appreciation to Universitas Satya Terra Bhinneka for providing financial support that enabled the completion of this research. The authors also gratefully acknowledge the Institute for Research and Community Service (LPPM) of Universitas Satya Terra Bhinneka for their continuous guidance, administrative assistance, and support throughout the research process. Their contributions have been instrumental in facilitating the successful execution of this study.

REFERENCES

[1] M. D. Sinaga, Fujiati, and D. P. F. Halawa, "Designing a Stunting Prediction Model Using Machine Learning to Support SDGs Achievement in Indonesia," *Sink. J. dan Penelit. Tek. Inform.*, vol. 9, no. 4, pp. 2070–2079, 2025, doi: https://doi.org/10.33395/sinkron.v9i4.15296 e-ISSN.

[2] I. Sutedja, L. Handayani, Maryuni, Hendry, and M. R. Faisal, "Analysis Implementation of Chatbots to Increase Knowledge of Stunting Prevention in Indonesian Society with Bibliometric," in *International Conference on Information Management and Technology (ICIMTech)*, 2024, pp. 473–477. doi: 10.1109/ICIMTech63123.2024.10780895.

[3] A. B. Zemariam, B. B. Abate, A. W. Alamaw, E. S. Lake, G. Yilak, M. Ayele, B. D. Tilahun, H. S. Ngusie, "Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine learning algorithms," *PLOS*, vol. 20, no. 1, pp. 1–13, 2025, doi: https://doi.org/10.1371/journal.pone.0316452.

[4] A. A. Permana, B. Raharja, and A. T. Perdana, "Artificial Intelligence for Diagnosing Child Stunting: A Systematic Literature Review," *J. Syst. Manag. Sci.*, vol. 13, no. 6, pp. 605–621, 2023, doi: 10.33168/JSMS.2023.0635.

[5] R. N. Utami, S. L. Panduragan, and N. Nambiar, "Leveraging Mobile Applications for Stunting Prevention in Indonesia : A Scoping Review," *MALAYSIAN J. Nurs.*, vol. 16, no. April, pp. 158–167, 2025, doi: 10.31674/mjn.2025.v16isupp2.018.

[6] A. Rusdianti, H. Santoso, W. Nugroho, B. J. Santosa, and Senarto, "Evaluating Acceleration of Stunting Prevention in Indonesia (2018-2024): A Roadmap-Based Program Analysis," *Heal. Dyn.*, vol. 2, no. 2, pp. 199–203, 2025, doi: https://doi.org/10.33846/hd20504 eISSN:

[7] F. T. Sabilillah, C. A. Sari, R. B. Abiyyi, and A. Susanto, "Comparison Of Machine Learning Algorithms On Stunting Detection For 'Centing' Mobile Application To Prevent Stunting," *Sinkron*, vol. 8, no. 4, pp. 2360–2368, 2024, doi: 10.33395/sinkron.v8i4.13967.

[8] G. S. Azahra and M. D. Kartikasari, "Child Stunting Classification using the LightGBM Method: A Case Study in the Rowosari District of Kendal, Central Java," *J. Math. Comput. Stat.*, vol. 8, no. 1, pp. 102–113, 2025, doi: 10.35580/jmathcos.v8i1.6865.

[9] A. Hendy, S. M. F Abdelaliem, A. Zaher, S. A. Alkubati, R. G. A. El-kader and A. Hendy , "Supervised machine learning for classification and prediction of stunting among under-five Egyptian children," *BMC Pediatr.*, vol. 25, no. 1, pp. 681–695, 2025, doi: 10.1186/s12887-025-06138-x.

[10] V. E. Togatorop, L. Rahayuwati, R. D. Susanti, and J. Y. Tan, "Stunting predictors among children aged 0-24 months in Southeast Asia: a scoping review," *Rev Bras Enferm*, vol. 77, no. 2, pp. 1–13, 2024, doi:

https://doi.org/10.1590/0034-7167-2022-0625.

[11] B. Rao, M. Rashid, M. G. Hasan, and G. Thunga, "Machine Learning in Predicting Child Malnutrition: A Meta-Analysis of Demographic and Health Surveys Data," *Int. J. Environ. Res. Public Health*, vol. 22, no. 3, pp. 1–15, 2025, doi: 10.3390/ijerph22030449.

[12] 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, pp. 89–121, 2024, doi: 10.3390/informatics11040089.

[13] E. Yogiyanti, Y. Yamasari, and E. Yohannes, "Analysis of the Application of Machine Learning Algorithms for Classification of Toddler Nutritional Status Based on Antropometric Data," *Indones. J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 4, pp. 594–603, 2025, doi: https://doi.org/10.35882/ijeeemi.v7i4.110.

[14] S. Banerjee and P. Shirisha, "Exploring the paradox of Muslim advantage in undernutrition among under - 5 children in India: a decomposition analysis," *BMC Pediatr.*, vol. 23, pp. 1–16, 2023, doi: 10.1186/s12887-023-04345-y.

[15] R. Gustriansyah, N. Suhandi, S. Puspasari, and A. Sanmorino, "Machine Learning Method to Predict the Toddlers' Nutritional Status," *J. INFOTEL*, vol. 16, no. 1, pp. 32–43, 2024, doi: 10.20895/INFOTEL.V15I4.988.

[16] C. J. Haug and J. M. Drazen, "Artificial Intelligence and Machine Learning in Clinical Medicine, 2023," *N. Engl. J. Med.*, vol. 388, no. 13, pp. 1201–1208, 2023, doi: 10.1056/nejmra2302038.

[17] H. Pallathadka, M. Mustafa, D. T. Sanchez, G. Sekhar Sajja, S. Gour, and M. Naved, "IMPACT OF MACHINE learning ON Management, healthcare AND AGRICULTURE," *Mater. Today Proc.*, vol. 80, no. xxxx, pp. 2803–2806, 2023, doi: 10.1016/j.matpr.2021.07.042.

[18] B. Ogwel, A. O. Awuor, C. Okonji, R.O. Anyango, C. Oreso, J. B. Ochieng, S. Munga, D. Nasrin, K. D. Tickell, P. B. Pavlinac, K. L. Kotloff, and R. Omore, "Predictive modelling of linear growth faltering among pediatric patients with Diarrhea in Rural Western Kenya: an explainable machine learning approach," *BMC Med. Inform. Decis. Mak.*, vol. 7, no. 2024, pp. 368–382, 2024, doi: 10.1186/s12911-024-02779-7.

[19] G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F. Amenta, "Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis," *J. Pers. Med.*, vol. 10, no. 2, pp. 1–11, 2020, doi: https://doi.org/10.3390/jpm10020021.

[20] B. Dou, Z. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jiang, Y. Zhu, J. Liu, B. Zhang, and G. Wei Wei, "Machine Learning Methods for Small Data Challenges in Molecular Science," *Chem. Rev.*, vol. 123, no. 13, pp. 8736–8780, 2023, doi: 10.1021/acs.chemrev.3c00189.

[21] R. Vogel and B. Mück, "Arti fi cial Intelligence — What to Expect From Machine Learning and Deep Learning in Hernia Surgery," *J. Abdom. Wall Surg.*, vol. 3, no. September, pp. 1–5, 2024, doi: 10.3389/jaws.2024.13059.

[22] G. Barakat, S. El, H. Hassan, N. Duong-trung, and W. Ramadan, "Enhancing Healthcare: Machine Learning for Diabetes Prediction and Retinopathy Risk Evaluation," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 7, pp. 18–36, 2024, doi: 10.14569/IJACSA.2024.0150703.

[23] R. H. Jhaveri, A. Revathi, K. Ramana, R. Raut, and R. K. Dhanaraj, "A Review on Machine Learning Strategies for Real-World Engineering Applications," *Hindawi*, vol. 2022, no. 1, pp. 1–26, 2022, doi: https://doi.org/10.1155/2022/1833507.

[24] H. Duan, "Personalized Self-Guided Tour Strategy by Integrating Random Forest Preference Model and Attractions Association Model," *Informatica*, vol. 49, no. 2025, pp. 1–18, 2025, doi: https://doi.org/10.31449/inf.v49i5.6002.

[25] O. Ojajuni, F. Ayeni, O. Akodu, F. Ekanoye, S. Adewole,T. Ayo, S. Misra, and V. Mbarika, "Predicting Student Academic Performance Using Machine Learning," *ICCSA*, vol. 12957. Spr, no. 1, pp. 481–491, 2021, doi: 10.1007/978-3-030-87013-3.

[26] F. U. Kaita, "PREDICTING UNDERNUTRITION RISK FACTORS USING MACHINE LEARNING TECHNIQUES IN NIGERIAN UNDER FIVE CHILDREN," *IJCSMC*, vol. 13, no. 7, pp. 56–70, 2024, doi: https://doi.org/10.47760/ijcsmc.2024.v13i07.006 Abstract—.

[27] A. V. Konstantinov and L. V. Utkin, "Interpretable machine learning with an ensemble of gradient boosting machines," *Knowledge-Based Syst.*, vol. 222, no. 1, pp. 1–28, 2021, doi: 10.1016/j.knosys.2021.106993.

[28] S. D. Kebede, Y. Sebastian, A. Yeneneh, and A. F. Chanie, "Prediction of contraceptive discontinuation among reproductive - age women in Ethiopia using Ethiopian Demographic and Health Survey 2016 Dataset : A Machine Learning Approach," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 9, pp. 1–17, 2023, doi: 10.1186/s12911-023-02102-w.

[29] H. Shen, H. Zhao, and Y. Jian, "Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea," *Children*, vol. 10, no. 1638, pp. 1–18, 2023, doi: https://doi.org/10.3390/children10101638.

[30] N. S. B. Sembiring, M. D. Sinaga, E. Ginting, F. Tahel, M. Fauzi, and Yudi, "Predict the Timeliness of Customer Credit Payments at Finance Companies Using a Decision Tree Algorithm," in *9th International Conference on Cyber and IT Service Management (CITSM)*, 2021, pp. 1–4. doi: 10.1109/CITSM52892.2021.9588880.

[31] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, "Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning

Techniques," *J. Prev. Med. Public Heal.*, vol. 56, no. 2023, pp. 41–49, 2023, doi: https://doi.org/10.3961/jpmph.22.388.

[32] A. Haque, N. Choudhury, B. Z. Wahid, SM. T. Ahmed, F. D. Farzana, M. Ali, F. Naz, T. J. Siddiqua, S. S. Rahman, A. S. G. Faruque and T. Ahmed, "A predictive modelling approach to illustrate factors correlating with stunting among children aged 12 – 23 months : a cluster randomised pre- - post study," *BMJ Open*, vol. 13, pp. 1–15, 2023, doi: 10.1136/bmjopen-2022-067961.

[33] F. Widyawati, H. P. Suhito, W. Yassin, and H. A. Santoso, "Classification of Toddler Nutritional Status Using Support Vector Machine and Random Forest Techniques With Optimal Feature Selection," *J. Tek. Inform.*, vol. 5, no. 6, pp. 1893–1904, 2025, doi: 10.52436/1.jutif.2024.5.6.4162.