Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparative Analysis of Automated Machine Learning Methods for Multiclass Stunting Prediction Using Anthropometric Data

Joharini[1)*], Agus Subekti[2)]
[1,2)]Faculty of Information Technology, Nusa Mandiri University, Jakarta, Indonesia
[1)]joharini@gmail.com, [2)]agus@nusamandiri.ac.id

Submitted :Jan 6, 2026 | **Accepted** : Feb 2, 2026 | **Published** : April 2, 2026

Abstract: Stunting remains a major public health challenge among children under five years old and requires reliable early screening to support timely nutritional interventions, particularly in resource-limited healthcare settings. However, many existing stunting prediction studies rely on complex socio-economic variables and manually selected machine learning models, which limits reproducibility and practical deployment. This study proposes an automated machine learning (AutoML)–based framework for multiclass stunting prediction using routinely collected anthropometric data. The prediction task is formulated as a multiclass classification problem encompassing normal growth, stunted, severely stunted, and above-normal nutritional status. The proposed framework integrates standardized preprocessing, systematic model comparison, stratified 10-fold cross-validation, and controlled hyperparameter optimization, evaluated under SMOTE and non-SMOTE preprocessing scenarios. Experimental results demonstrate that reliable multiclass prediction can be achieved without socio-economic variables. Under SMOTE preprocessing, the optimized k-Nearest Neighbors model improves minority-class sensitivity, increasing accuracy from 0.9806 to 0.9820 with an MCC of 0.9688, while under non-SMOTE conditions, Random Forest achieves robust performance with an accuracy of 0.9985 and an MCC of 0.9975 without resampling. Confusion matrix, ROC, and learning curve analyses confirm strong discriminative capability and stable generalization for both models. Overall, the findings indicate that the proposed AutoML-based framework provides a practical, scalable, and reproducible solution for early multiclass stunting screening using anthropometric data alone.

Keywords: Anthropometric Data, Automated Machine Learning, Multiclass Classification, SMOTE, Stunting Prediction

INTRODUCTION

Stunting remains a major global public health concern, particularly among children under five years old in low- and middle-income countries. As a manifestation of chronic malnutrition, it adversely affects linear growth, cognitive development, and long-term health outcomes, with lasting consequences for education, productivity, and non-communicable disease risk, thereby imposing a substantial burden on health systems and socio-economic development (Ayele et al., 2025), (Rao et al., 2025), (Zemariam et al., 2025).

Recent studies have increasingly adopted data-driven approaches for early identification and risk assessment of childhood stunting. Although conventional epidemiological and statistical methods have contributed to identifying key determinants, they often struggle to model complex and nonlinear interactions among biological, environmental, and socio-demographic factors. Consequently, machine learning (ML) techniques have emerged as promising alternatives due to their ability to process high-dimensional data and uncover latent patterns (Ndagijimana et al., 2023), (Shen et al., 2023), (Shirin Sara et al., 2024).

Various ML algorithms, including Random Forest, Support Vector Machine, Gradient Boosting, and ensemble-based methods, have demonstrated encouraging predictive performance across diverse geographical contexts such as Ethiopia, Rwanda, Indonesia, and Papua New Guinea, (Anggito Herlambang Hadisuwarno et al., 2025), (Hasdyna et al., 2024), (Hendy et al., 2025), (Elim & Utami, 2025). More recently, deep learning approaches have been explored for longitudinal nutritional data, further highlighting the potential of advanced learning architectures in stunting prediction (Begashaw et al., 2025), (Lestari et al., 2024).

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Despite these advances, practical deployment of ML-based stunting prediction remains limited. Many studies rely on socio-economic variables such as household income, parental education, sanitation, and dietary diversity which are often unavailable or inconsistently recorded in routine primary healthcare settings, restricting scalability and real-world applicability (Zemariam et al., 2025), (Tamanna et al., 2025), (Wicaksono & Harsanti, 2020). In addition, most studies formulate stunting prediction as a binary task, which oversimplifies clinical practice where nutritional status is inherently multiclass, including normal, stunted, severely stunted, and above-normal growth (Lestari et al., 2024), (Ratnasari et al., 2024).

A review of the literature reveals several gaps. First, although ML has been widely applied, relatively few studies emphasize anthropometric-only data for early stunting screening, limiting applicability in low-resource and time-constrained healthcare settings (Ratnasari et al., 2024), (Husaini et al., 2023), (Sutarmi et al., 2023).

Second, systematic and reproducible model selection remains underexplored, as many studies rely on manually selected algorithms or limited comparisons, introducing potential bias and reducing reproducibility (Anggito Herlambang Hadisuwarno et al., 2025), (Putri et al., 2024), (Sinaga et al., 2025).

Third, while ensemble and hybrid methods have been proposed (Hasdyna et al., 2024), (Husaini et al., 2023), (Heryati et al., 2025), the application of automated machine learning (AutoML) frameworks for multiclass stunting prediction is still limited, with insufficient emphasis on transparency and generalizability (Indrisari et al., 2025), (Mulyani et al., 2025).

Finally, few studies explicitly frame stunting prediction as a practical early screening tool rather than a purely analytical model, highlighting the need for solutions that balance predictive accuracy with simplicity, interpretability, and deployability in primary healthcare systems.

Based on the identified gaps, this study aims to develop a multiclass stunting prediction model using minimal anthropometric indicators of children under five years old. The study applies an automated machine learning framework to enable systematic and unbiased model comparison across multiple supervised classification algorithms. Model performance is evaluated using standardized metrics to ensure methodological rigor and comparability. Furthermore, the study assesses the feasibility of deploying the proposed model as an early screening tool in resource-limited primary healthcare settings.

This study contributes to the literature by demonstrating that effective multiclass stunting prediction can be achieved using minimal anthropometric measurements without reliance on socio-economic variables. By employing an AutoML framework, the study reduces model selection bias and enhances reproducibility through systematic experimentation. The prediction framework aligns outputs with clinically relevant multiclass nutritional categories, thereby improving practical interpretability. Additionally, the study emphasizes scalability and real-world deployability to support early detection initiatives in community and primary healthcare systems.

The remainder of this paper is structured as follows. Section 2 reviews related work on stunting prediction and machine learning approaches. Section 3 describes the dataset, preprocessing procedures, AutoML framework, and experimental design. Section 4 presents and discusses the experimental results. Finally, Section 5 concludes the study, outlining its limitations, future research directions, and practical implications.

LITERATURE REVIEW

Stunting and Child Malnutrition Studies

Stunting is widely recognized as a key indicator of chronic malnutrition in early childhood and remains a major public health concern in low- and middle-income countries. Global reports by UNICEF, the World Health Organization (WHO), and the World Bank consistently document its persistent prevalence and long-term impacts on physical growth, cognitive development, and health outcomes across the life course (UNICEF & others, 2016). These reports emphasize that stunting is a multidimensional condition influenced by biological, nutritional, environmental, and socio-economic factors.

Extensive epidemiological studies using cross-sectional and longitudinal data show that inadequate early-life nutrition significantly impairs linear growth and long-term health trajectories (Zemariam et al., 2025), (Tamanna et al., 2025), (Wicaksono & Harsanti, 2020). Multilevel analyses further identify household economic status, parental education, sanitation, and healthcare access as major contributors to stunting prevalence (Tamanna et al., 2025), (Wicaksono & Harsanti, 2020). While these studies provide important causal and policy-relevant insights, their explanatory focus limits direct applicability for predictive early screening.

Recent work has also explored biological mechanisms, such as the association between gut microbiome composition and stunting (Chibuye et al., 2024). Although biologically informative, such approaches require specialized measurements and complex analyses, constraining scalability in routine primary healthcare. Overall, existing studies highlight the importance of early identification but reveal a gap in predictive, scalable, and data-efficient screening approaches using routinely available data.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Machine Learning for Stunting Prediction

To address the limitations of conventional statistical methods, machine learning (ML) techniques have increasingly been applied to stunting prediction due to their ability to model nonlinear relationships among multiple predictors. Supervised algorithms such as Logistic Regression, Support Vector Machine (SVM), Random Forest, and Gradient Boosting have been widely investigated (Ndagijimana et al., 2023), (Shen et al., 2023), (Shirin Sara et al., 2024).

Empirical studies across diverse regions report encouraging performance. In Papua New Guinea, ML models outperformed traditional statistical approaches in predicting stunting among under-five children (Shen et al., 2023). Similar findings were reported in Rwanda and Ethiopia, where tree-based and ensemble models achieved superior accuracy (Ayele et al., 2025), (Ndagijimana et al., 2023).

In Indonesia, ML-based stunting prediction has gained increasing attention, with ensemble methods such as XGBoost and Random Forest consistently demonstrating strong performance on anthropometric and health-related data (Anggito Herlambang Hadisuwarno et al., 2025). (Elim & Utami, 2025), (Sutarmi et al., 2023), (Putri et al., 2024), (Sinaga et al., 2025). However, most studies rely on manually selected algorithms and heterogeneous experimental designs, limiting reproducibility and cross-study comparability.

Deep Learning and Multiclass Stunting Classification

Beyond conventional ML, deep learning approaches particularly recurrent neural networks and long short-term memory (LSTM) models have been explored for longitudinal stunting prediction, demonstrating improved performance by capturing temporal nutritional patterns (Begashaw et al., 2025). Several studies emphasize the importance of framing stunting prediction as a multiclass problem to better reflect clinical nutritional assessment, distinguishing normal, stunted, and severely stunted (Lestari et al., 2024), (Ratnasari et al., 2024).

Despite their potential, deep learning models typically require large datasets, substantial computational resources, and extensive parameter tuning, limiting their practicality for early screening in low-resource settings. Recent reviews indicate that automated and standardized model selection frameworks remain underutilized in stunting prediction studies, particularly in Indonesia (Indrisari et al., 2025), (Purwati & Widiantoro, 2025). Comparative studies further highlight the need for consistent preprocessing, cross-validation, and standardized evaluation to ensure transparent and reproducible model assessment (Anggito Herlambang Hadisuwarno et al., 2025), (Putri et al., 2024), (Sinaga et al., 2025). In this context, automated machine learning (AutoML) frameworks offer a promising solution for unbiased model comparison and improved reproducibility.

Comparative Analysis of Previous Studies

To clarify the positioning of this research, Table 1 summarizes representative ML-based stunting prediction studies, highlighting differences in data sources, feature complexity, classification strategies, and methodological limitations.

Table 1. Comparative Summary of Previous Stunting Prediction Studies

Study	Country / Data Source	Features Used	Classification Type	Methodology	Key Limitations
Ayele et al. (Ayele et al., 2025)	Ethiopia	Anthropometric + socio-economic	Binary	Ensemble ML	High feature dependency
Rao et al. (Rao et al., 2025)	Multi-country DHS	Demographic & health variables	Binary	Meta-analytical ML	Not screening-oriented
Zemariam et al. (Zemariam et al., 2025)	Ethiopia	Socio-economic + anthropometric	Binary	ML classifiers	Focus on adolescents
Ndagijimana et al (Ndagijimana et al., 2023)	Rwanda	Anthropometric + health	Binary	ML models	Manual model selection
Shen et al. (Shen et al., 2023)	Papua New Guinea	Anthropometric	Binary	ML algorithms	Limited class granularity
Hadisuwarno et al. (Anggito Herlambang Hadisuwarno et al., 2025)	Indonesia	Anthropometric + health	Binary	XGBoost, RF	No AutoML
Hasdyna et al. (Hasdyna et al., 2024)	Indonesia	Mixed variables	Binary	Hybrid ML	Complex pipeline
Begashaw et al. (Begashaw et al., 2025)	Ethiopia	Longitudinal nutrition	Multiclass	Deep Learning (LSTM)	High computational cost

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Study	Country / Data Source	Features Used	Classification Type	Methodology	Key Limitations
Lestari et al. (Lestari et al., 2024)	Indonesia	Anthropometric	Multiclass	Deep Neural Network	Manual tuning
Ratnasari et al. (Ratnasari et al., 2024)	Indonesia	Anthropometric	Multiclass	ML classifiers	No systematic comparison
This Study	Under-five children	**Anthropometric-only**	**Multiclass**	**AutoML-based ML**	Single dataset

Research Positioning and Contribution of This Study

Based on the reviewed literature, this study is positioned at the intersection of anthropometric-based stunting prediction, multiclass classification, and automated model selection. Unlike prior works that rely on extensive socio-economic variables, complex pipelines, or manually selected algorithms, this study emphasizes minimal anthropometric indicators and employs an AutoML framework to ensure objective, systematic, and reproducible evaluation.

By explicitly targeting multiclass early screening applicability, this research addresses key gaps identified in previous studies and provides a coherent foundation for the experimental design in Section 3, with comparative results and implications discussed in Sections 4 and 5.

METHOD

Research Design

This study adopts a quantitative research design based on supervised machine learning to develop a multiclass prediction model for childhood stunting among children under five years old. Nutritional status prediction is formulated as a multiclass classification task using anthropometric measurements routinely collected in primary healthcare services.

To ensure methodological rigor and reproducibility, an Automated Machine Learning (AutoML) framework is employed. AutoML standardizes preprocessing, enables systematic comparison of multiple classifiers, and minimizes subjectivity in model selection, directly addressing issues of reproducibility, bias, and practical applicability for early screening.

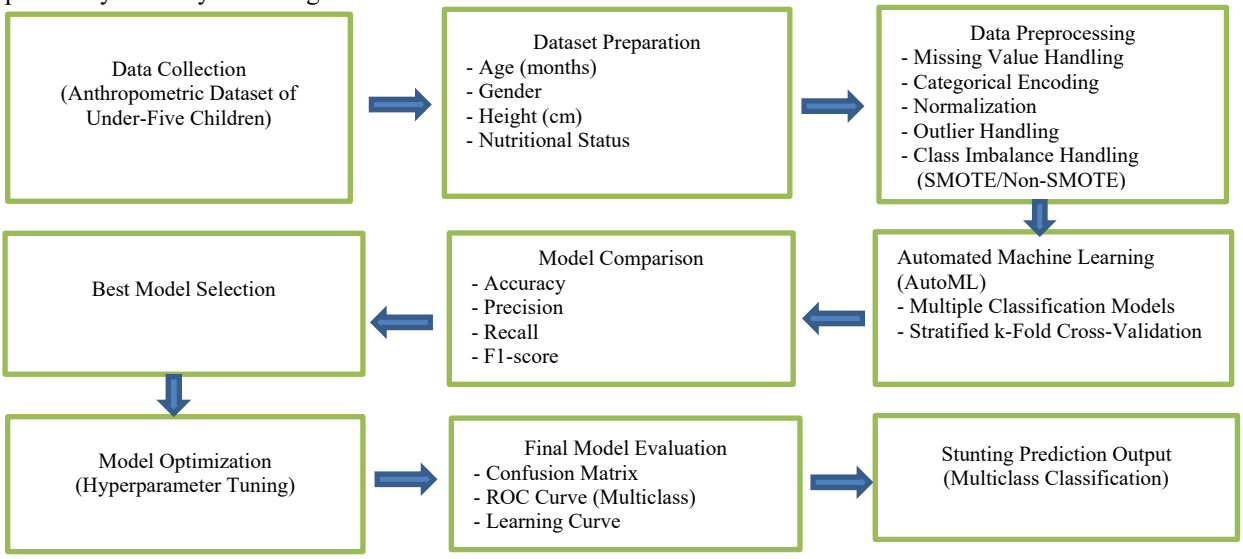

Figure 1. Workflow of the proposed AutoML-based multiclass stunting prediction framework.

Figure 1 illustrates the overall workflow, encompassing data acquisition, preprocessing under SMOTE and non-SMOTE scenarios, automated model comparison using stratified k-fold cross-validation, best model selection, hyperparameter optimization, and final multiclass prediction.

Dataset Description

The dataset is described in the third paragraph of this section to ensure logical sequencing from research design to data specification. The dataset consists of 120,999 anthropometric records of children under five years old, where each record represents an individual child. The predictor variables include age in months, gender, and height in centimeters, while nutritional status serves as the target variable.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Nutritional status is defined as a multiclass label comprising four categories: normal growth, stunted, severely stunted, and above-normal growth. This categorization aligns with standard clinical nutritional assessment practices and provides greater granularity than binary formulations.

The nutritional status labels are determined according to established WHO height-for-age standards. To prevent data leakage, only raw anthropometric measurements are used as input features, and no derived Z-score variables are included in the modeling process. All predictor variables are routinely recorded in primary healthcare services, supporting the feasibility of real-world implementation as an early screening tool. Dataset attributes are summarized in Table 2.

Table 2. Dataset Attributes and Descriptions

Attribute Name	Data Type	Description
Age (months)	Numerical	Age of the child measured in months
Gender	Categorical	Biological sex of the child (male/female)
Height (cm)	Numerical	Measured height of the child in centimeters
Nutritional Status	Categorical (Target)	Child nutritional status: normal, stunted, severely stunted, above-normal

Data Preprocessing

All preprocessing steps are implemented within a unified AutoML pipeline to ensure consistent treatment across models. The preprocessing procedure includes missing value handling, categorical encoding for gender, feature scaling for numerical variables, and outlier mitigation.

To assess the effect of class imbalance on predictive performance, two preprocessing scenarios are evaluated. In the first scenario, the Synthetic Minority Over-sampling Technique (SMOTE) is applied exclusively to the training folds to balance minority classes. In the second scenario, models are trained using the original class distribution without oversampling. Both scenarios apply identical preprocessing steps and classification models, enabling fair and objective performance comparison. Figure 2 illustrates the dual preprocessing strategy.

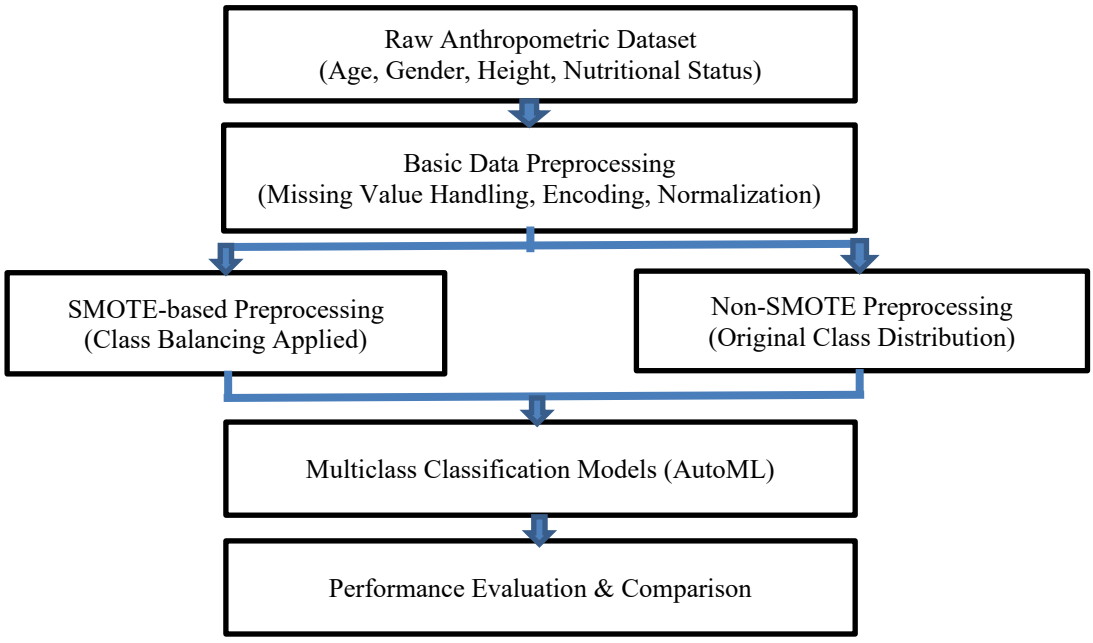

Figure 2. Data preprocessing scenarios for stunting prediction.

Automated Machine Learning Framework

The AutoML framework integrates preprocessing, model training, stratified k-fold cross-validation, hyperparameter optimization, and performance evaluation within a unified pipeline. Multiple supervised classifiers representing different modeling paradigms are evaluated, including Logistic Regression, k-Nearest Neighbors, Decision Tree, Random Forest, and Gradient Boosting–based models.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Stratified k-fold cross-validation is employed to preserve class distributions across folds, which is essential for reliable multiclass performance estimation. Model selection is based on a multi-metric evaluation strategy, including accuracy, precision, recall, F1-score, Cohen's Kappa, and Matthews Correlation Coefficient, ensuring balanced assessment for early screening applications.

Model Training and Optimization

After identifying the most suitable classifier through automated comparison, hyperparameter optimization is performed to further improve predictive performance (Raschka, 2020), (Platos, 2025). Optimization is conducted under identical cross-validation settings to maintain experimental consistency.

The optimized model is subsequently retrained using the complete training dataset to fully exploit available information (Platos, 2025). Final evaluation on unseen data assesses generalization capability and mitigates overfitting, forming the basis for the analysis presented in Chapter IV.

Model Evaluation Metrics

The performance of the proposed multiclass stunting prediction models is evaluated using several standard classification metrics derived from the confusion matrix (Raschka, 2020), (Sathyanarayanan & Tantri, 2024). For a multiclass classification problem with K nutritional status categories, the confusion matrix is represented as:

$$CM = \begin{bmatrix} c11 & c12 & \cdots & c1K \\ c21 & c22 & \dots & c2K \\ \vdots & \vdots & \ddots & \vdots \\ cK1 & cK2 & \dots & cKK \end{bmatrix} \quad (1)$$

where c_{ij} denotes the number of instances belonging to class i that are predicted as class j. For a multiclass classification problem with K classes, class-specific true positives TP_k, false positives FP_k, and false negatives FN_k are derived from the confusion matrix using a one-vs-rest (OvR) formulation (Platos, 2025), (Sathyanarayanan & Tantri, 2024), in which class k is treated as the positive class and all remaining classes are considered as negative.

Accuracy

Accuracy measures the overall proportion of correctly classified instances across all classes and is defined as (Raschka, 2020), (Platos, 2025):

$$Accuracy = \frac{\sum_{i=1}^{K} c_{ii}}{\sum_{i=1}^{K}\sum_{j=1}^{K} c_{ij}} \quad (2)$$

Precision

Precision reflects the correctness of positive predictions for a given class K and is calculated as:

$$Precision_k = \frac{TP_k}{TP_k + FP_k} \quad (3)$$

where TP_k represents the number of true positives and FP_k represents the number of false positives for class k.

Recall (Sensitivity)

Recall measures the ability of the model to correctly identify instances of a particular class (Raschka, 2020), (Cabot & Ross, 2024) and is given by:

$$Recall_k = \frac{TP_k}{TP_k + FN_k} \quad (4)$$

Where FN_k denotes the number of false negatives for class k. In the context of early stunting screening, recall is particularly important, as failing to identify children at risk of stunting may delay necessary nutritional interventions.

F1-Score

The F1-score represents the harmonic mean of precision and recall and provides a balanced measure of classification performance (Takahashi et al., 2022):

$$F1\text{-}score_k = \frac{2 \times Precision_k \times Recall_k}{Precision_k + Recall_k} \quad (5)$$

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Macro-Averaged Metrics

Given the multiclass nature of the stunting prediction problem, all evaluation metrics are aggregated using macro-averaging to ensure equal importance across nutritional status categories (Takahashi et al., 2022):

$$Macro\text{-}Metric = \frac{1}{K}\sum_{k=1}^{K} Metric_k \quad (6)$$

This approach prevents dominant classes from disproportionately influencing overall performance evaluation.

Area Under the ROC Curve (AUC)

The Area Under the Receiver Operating Characteristic Curve (AUC) is used to evaluate the model's discriminative ability (Cabot & Ross, 2024). For multiclass classification, the AUC is computed using a one-vs-rest strategy and is defined as:

$$AUC = \int_0^1 TPR(FPR)\ d(FPR) \quad (7)$$

Where TPR denotes the true positive rate and FPR denotes the false positive rate. In multiclass classification, the AUC is computed using a one-vs-rest (OvR) strategy, and the final AUC value is reported as the macro-averaged AUC across all classes.

Ethical Considerations

This study uses fully anonymized secondary data without personally identifiable information. As no direct human subject involvement is present, ethical risks related to privacy and confidentiality are minimal.

Methodological Link to Subsequent Chapters

The methodological framework described in this chapter establishes the foundation for the experimental evaluation and comparative analysis presented in the following section. The integration of AutoML, dual preprocessing strategies, and standardized evaluation metrics ensures methodological transparency, reproducibility, and alignment with the research objectives.

RESULTS

Experimental Setup Overview

This section presents the experimental results of the proposed AutoML-based multiclass stunting prediction framework using 120,999 anthropometric records of children under five years old. The dataset distribution comprised 56.0% normal growth, 11.4% stunted, 16.4% severely stunted, and 16.2% above-normal growth, indicating moderate class imbalance. To address this, both SMOTE and non-SMOTE preprocessing scenarios were evaluated. The AutoML framework was implemented in a structured Python-based environment integrating preprocessing, automated model comparison, stratified 10-fold cross-validation, hyperparameter optimization, and multi-metric evaluation within a unified pipeline. Multiple supervised classifiers, including Logistic Regression, k-Nearest Neighbors, Decision Tree, Random Forest, Extra Trees, and Gradient Boosting models, were systematically compared under identical validation settings.

Model performance was assessed using accuracy, precision, recall, F1-score, Cohen's Kappa, and Matthews Correlation Coefficient (MCC), with emphasis on F1-score and MCC for robustness in multiclass evaluation. Stratified cross-validation ensured stable and unbiased performance estimation across folds. Despite near-perfect baseline metrics for ensemble models, learning curve convergence and consistent agreement-based measures indicate stable generalization behavior without substantial evidence of severe overfitting. The results are presented from baseline comparison to optimized model evaluation following the methodological workflow outlined in the previous chapter.

A. Baseline Model Comparison (AutoML Output)

SMOTE Preprocessing Scenario

Table 3 summarizes baseline model performance under SMOTE preprocessing.

Table 3. Baseline Performance Comparison (SMOTE Scenario)

Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
K Neighbors Classifier	0.9806	0.9806	0.9811	0.9806	0.9687	0.9688
Decision Tree Classifier	0.9784	0.9784	0.9795	0.9785	0.9651	0.9654

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Model	Accuracy	Recall	Prec.	F1	Kappa	MCC
Extra Trees Classifier	0.9770	0.9770	0.9775	0.9770	0.9627	0.9630
Random Forest Classifier	0.9769	0.9769	0.9776	0.9769	0.9625	0.9628
Extreme Gradient Boosting	0.9656	0.9656	0.9670	0.9657	0.9444	0.9448

The k-Nearest Neighbors (KNN) classifier achieved the strongest baseline results, with an F1-score of 98.06% and an MCC of 96.88%, while maintaining comparable accuracy (98.06%). This indicates that distance-based learning benefits from balanced class distributions and standardized feature scaling. The strong KNN performance suggests that anthropometric features form clear local proximity structures when class imbalance is mitigated. However, sensitivity to synthetic samples highlights the need for careful interpretation in real-world deployment.

Baseline Performance under Non-SMOTE Preprocessing

Baseline results under the original class distribution are shown in Table 4.

Table 4. Baseline Performance Comparison (Non-SMOTE Scenario)

Model	Accuracy	Recall	Precision	F1-score	Kappa	MCC
Random Forest Classifier	0.9985	0.9985	0.9985	0.9985	0.9975	0.9975
Extra Trees Classifier	0.9985	0.9985	0.9985	0.99875	0.9975	0.9975
Decision Tree Classifier	0.9979	0.9979	0.9979	0.9979	0.9967	0.9967
K Neighbors Classifier	0.9958	0.9958	0.9958	0.9958	0.9933	0.9933
Extreme Gradient Boosting	0.9899	0.9899	0.9899	0.9899	0.9837	0.9837

Random Forest achieved near-perfect performance, with accuracy and F1-score of 99.85% and an MCC of 99.75%, demonstrating the effectiveness of ensemble-based tree models in capturing nonlinear relationships within anthropometric data. These results are interpreted as baseline references to contrast preprocessing effects rather than to determine final model selection. Compared to distance-based methods, Random Forest is less sensitive to feature scaling and class imbalance, supporting its robustness under naturally skewed data distributions.

B. Comparative Analysis of Selected Models

A direct comparison between the best-performing models under each preprocessing strategy is presented in Table 5.

Table 5. Comparative Performance of Selected Models

Metric	KNN (SMOTE)	Random Forest (Non-SMOTE)
Accuracy	0.9806	**0.9985**
F1-score	0.9806	**0.9985**
MCC	0.9688	**0.9975**
Model Characteristics	Distance-based	Ensemble-based
Sensitivity to Imbalance	High	Low

As shown in Table 5, the Random Forest model without resampling achieved superior global agreement metrics, with accuracy and F1-score of 99.85% and an MCC of 99.75%, whereas the KNN model under SMOTE attained an accuracy and F1-score of 98.06% and an MCC of 96.88%. While Random Forest demonstrates higher overall consistency under the natural class distribution, KNN shows improved sensitivity to minority classes when synthetic balancing is applied, indicating that preprocessing strategy and algorithm characteristics jointly influence performance.

Although the near-perfect metrics of Random Forest warrant consideration of potential overfitting, performance was estimated using stratified 10-fold cross-validation rather than a single train–test split, reducing optimistic bias. The alignment of agreement-based metrics with accuracy and the learning curve convergence presented in Table 5 further support stable generalization behavior. Consequently, the strong baseline performance of Random Forest is interpreted as robust model fit supported by rigorous validation, with its reduced sensitivity to feature scaling and class imbalance explaining its advantage under the naturally skewed distribution.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

C. Best Model Evaluation

Hyperparameter Tuning

Table 6 summarizes the hyperparameter tuning results for both models.

Model	Hyperparameter	Baseline Value	Optimized Value	Accuracy
KNN	(K)	5	**2**	**0.9820**
Random Forest	`max_depth`	None	5	0.8802

Hyperparameter optimization was applied to the selected models using identical cross-validation settings. For KNN, tuning focused on the number of nearest neighbors (K). Reducing K from 5 to 2 improved mean accuracy from 0.9806 to 0.9820, accompanied by consistent gains in recall, F1-score, Cohen's Kappa, and MCC (Table 6). This indicates that a smaller neighborhood size better captures local anthropometric patterns relevant to multiclass stunting classification.

In contrast, Random Forest tuning did not yield performance improvements. Constraining tree depth reduced mean accuracy to 0.8802, indicating underfitting due to limited representational capacity. These results confirm that the baseline Random Forest configuration was already well aligned with dataset complexity.

Confusion Matrix Analysis

Confusion matrix analysis shown in Figure 3 demonstrates strong diagonal dominance for both models, indicating high overall classification accuracy.

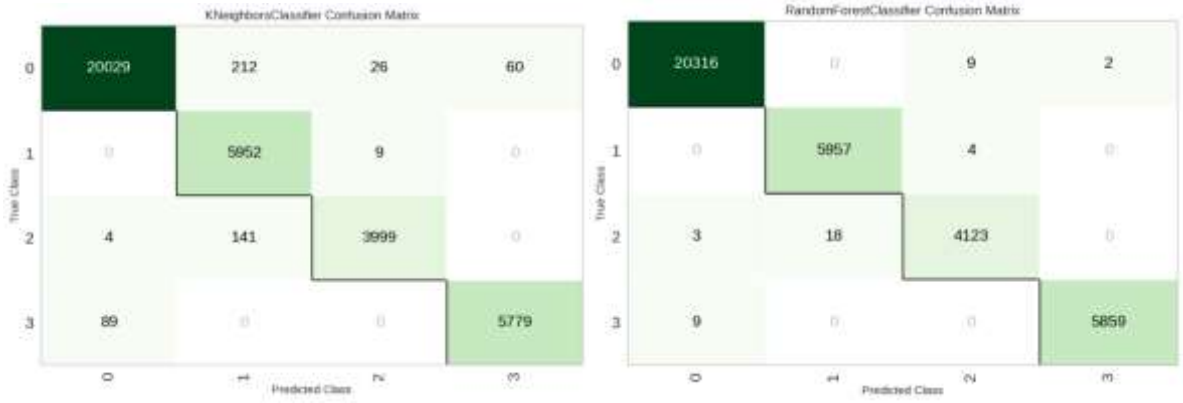

Figure 3. Confusion Matrix of the Tuned KNN and Random Forest Models

Random Forest exhibits highly stable class separation with minimal misclassification, whereas the tuned KNN shows slightly higher confusion between adjacent categories, particularly normal and stunted. Nevertheless, the tuned KNN maintains high recall for stunted and severely stunted classes, which is critical for early screening applications where false negatives carry greater public health risk.

ROC Curve Analysis

Receiver Operating Characteristic (ROC) curves were used to evaluate the discriminative ability of the tuned KNN and Random Forest models under a multiclass one-vs-rest setting, as illustrated in Figure 4.

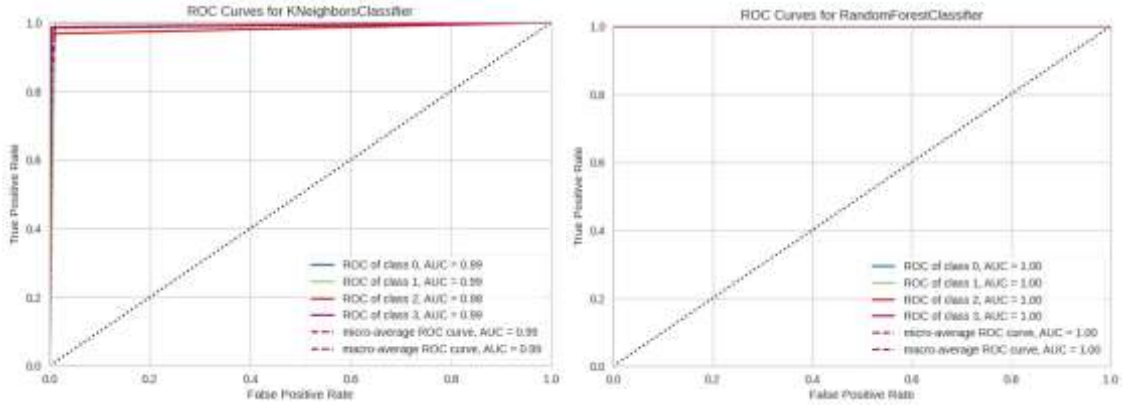

Figure 4. ROC Curves of the Tuned KNN and Random Forest Models

ROC analysis (Figure 4) confirms both models demonstrate macro-averaged AUC values approaching 0.99, confirming strong class separability. Although Random Forest exhibits marginally higher global separability, the

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 2, April 2026
DOI : https://doi.org/10.33395/sinkron.v10i2.15886
e-ISSN : 2541-2019
p-ISSN : 2541-044X

difference does not substantially alter screening utility. High AUC values across folds further support the stability of the models and reduce concerns regarding threshold-dependent overfitting.

Learning Curve Analysis

Learning curves were analyzed to assess the generalization behavior of the tuned KNN and Random Forest models, as shown in Figure 5.

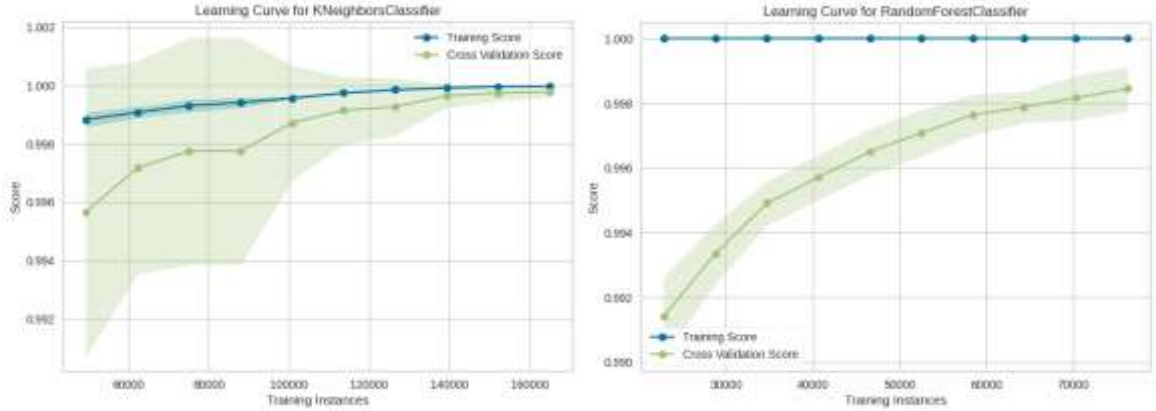

Figure 5. Learning Curves of the Tuned KNN and Random Forest Models

Learning curves shown in Figure 5 provide direct evidence regarding overfitting and generalization. Both models exhibit convergence between training and validation performance as sample size increases. Random Forest demonstrates stable performance with minimal gap between training and validation curves, indicating strong generalization capability despite high accuracy values. The tuned KNN model shows gradual improvement with increasing data size, suggesting that additional data enhances performance without inducing divergence between training and validation metrics.

The convergence pattern observed in Figure 5 indicates that the models do not suffer from severe overfitting. Instead, the large dataset size of over 120,000 records appears to support stable parameter estimation and reliable generalization.

DISCUSSION

The results demonstrate that systematic AutoML-based model comparison is essential for multiclass stunting prediction, as performance varies depending on preprocessing strategy and algorithm characteristics. Under SMOTE preprocessing, KNN achieved strong performance with enhanced sensitivity to minority nutritional categories, indicating that distance-based learning benefits from balanced class distributions. Under non-SMOTE conditions, Random Forest exhibited robust and highly consistent performance without the need for synthetic resampling, reflecting the strength of ensemble-based tree models in capturing nonlinear relationships within anthropometric features.

Hyperparameter optimization produced model-dependent effects, improving KNN through neighborhood refinement while confirming the adequacy of the baseline Random Forest configuration. Confusion matrix and ROC analyses indicate strong discriminative capability across nutritional categories. Learning curve convergence between training and validation performance further supports stable generalization behavior. Despite near-perfect baseline metrics, the absence of a significant gap between training and validation scores and the use of stratified 10-fold cross-validation reduce the likelihood of severe overfitting. The variance-reducing mechanism of Random Forest ensembles also contributes to its stable performance across folds.

Overall, these findings confirm that effective multiclass stunting prediction can be achieved using anthropometric data alone. The integration of a structured AutoML framework, dual preprocessing strategies, and comprehensive cross-validation enhances methodological transparency and reproducibility while maintaining practical relevance for early screening in primary healthcare settings.

CONCLUSION

This study developed an automated machine learning framework for multiclass stunting prediction using routinely collected anthropometric data from children under five years old. Through systematic model comparison and stratified cross-validation within a unified AutoML pipeline, the proposed approach improves methodological rigor and reproducibility in public health prediction tasks. The findings demonstrate that reliable multiclass classification can be achieved without reliance on complex socio-economic variables, while maintaining stable generalization performance.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

The comparative analysis shows that ensemble-based models provide strong robustness under naturally imbalanced distributions, whereas distance-based methods benefit from resampling strategies when minority sensitivity is prioritized. Overall, the framework offers a scalable and practically deployable solution for early stunting screening in resource-limited healthcare settings.

Future research should validate the framework across multi-site or nationally representative datasets to strengthen external validity. Incorporating longitudinal growth measurements and explainable artificial intelligence techniques may further enhance predictive insight and real-world implementation within healthcare systems.

REFERENCES

Anggito Herlambang Hadisuwarno, M., Teguh Martono, K., Adriono, E., Soedarto, J., Tembalang, K., Semarang, K., Tengah, J., & Artikel, R. (2025). Komparasi performa model machine learning algoritma XGBoost dan Random Forest pada studi kasus mendeteksi stunting. *AITI: Jurnal Teknologi Informasi*, *22*(2), 266–278.

Ayele, M. K., Baye, G. A., Yesuf, S. H., Engda, A. A., & Mitiku, E. T. (2025). Predicting stunting status among under five children in ethiopia using ensemblemachine learning algorithms. *Scientific Reports*, *15*(1), 1–11. https://doi.org/10.1038/s41598-025-03206-1

Begashaw, G. B., Zewotir, T., & Fenta, H. M. (2025). A deep learning approach for classifying and predicting children's nutritional status in Ethiopia using LSTM-FC neural networks. *BioData Mining*, *18*(1). https://doi.org/10.1186/s13040-025-00425-0

Cabot, J. H., & Ross, E. G. (2024). *Evaluating Prediction Model Performance*. *174*(3), 723–726. https://doi.org/10.1016/j.surg.2023.05.023.Evaluating

Chibuye, M., Mende, D. R., Spijker, R., Simuyandi, M., Luchen, C. C., Bosomprah, S., Chilengi, R., Schultsz, C., & Harris, V. C. (2024). Systematic review of associations between gut microbiome composition and stunting in under-five children. *Npj Biofilms and Microbiomes*, *10*(1). https://doi.org/10.1038/s41522-024-00517-5

Elim, M. I., & Utami, E. (2025). Performance Comparison of Child Stunting Prediction Support Vector Machine vs Random Forest with Grid Search Optimization. *Jurnal Teknik Informatika (Jutif)*, *6*(5), 5305–5319. https://doi.org/10.52436/1.jutif.2025.6.5.5285

Hasdyna, N., Dinata, R. K., Rahmi, & Fajri, T. I. (2024). Hybrid Machine Learning for Stunting Prevalence: A Novel Comprehensive Approach to Its Classification, Prediction, and Clustering Optimization in Aceh, Indonesia. *Informatics*, *11*(4). https://doi.org/10.3390/informatics11040089

Hendy, A., Ibrahim, R. K., Abdelaliem, S. M. F., Zaher, A., Alkubati, S. A., El-kader, R. G. A., & Hendy, A. (2025). Supervised machine learning for classification and prediction of stunting among under-five Egyptian children. *BMC Pediatrics*, *25*(1). https://doi.org/10.1186/s12887-025-06138-x

Heryati, A., Marcelina, D., & Romli, H. (2025). Optimization of Stunting Risk Prediction Using a Hybrid Genetic-Machine Learning Model. *Journal of Artificial Intelligence and Software Engineering*, *5*(2), 807–815. https://doi.org/10.30811/jaise.v5i2.6988

Husaini, A., Hoeronis, I., Lumana, H. H., & Puspareni, L. D. (2023). Early Detection of Stunting in Toddlers Based on Ensemble Machine Learning in Purbaratu Tasikmalaya. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, *11*(3), 487. https://doi.org/10.26418/justin.v11i3.66465

Indrisari, E., Febiansyah, H., & Adiwinoto, B. (2025). A Systematic Literature Review on the Application of Machine Learning for Predicting Stunting Prevalence in Indonesia (2020–2024). *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *14*(3), 277–283. https://doi.org/10.32736/sisfokom.v14i3.2366

Lestari, W. S., Saragih, Y. M., & Caroline, C. (2024). Multiclass Classification for Stunting Prediction Using Deep Neural Networks. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *10*(2), 386–393. https://doi.org/10.33480/jitk.v10i2.5636

Mulyani, H., Faturrochman, R., & Permana, D. H. (2025). *Machine Learning-Based Early Detection of Stunting and Intervention Recommendations*. *8*(2). https://doi.org/10.32877/bt.v8i2.3213

Ndagijimana, S., Kabano, I. H., Masabo, E., & Ntaganda, J. M. (2023). Prediction of Stunting among Under-5 Children in Rwanda Using Machine Learning Techniques. *Journal of Preventive Medicine and Public Health*, *56*(1), 41–49. https://doi.org/10.3961/jpmph.22.388

Platos, J. (2025). *Fundamentals of Machine Learning*.

Purwati, N., & Widiantoro, T. (2025). AI and Machine Learning untuk Diagnosis dan Intervensi Dini pada Stunting Balita: A Systematic Literature Review. *Infomatek*, *27*(1), 71–86. https://doi.org/10.23969/infomatek.v27i1.24136

Putri, I. P., Terttiaavini, T., & Arminarahmah, N. (2024). Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 257–265. https://doi.org/10.57152/malcom.v4i1.1078

Rao, B., Rashid, M., Hasan, M. G., & Thunga, G. (2025). Machine Learning in Predicting Child Malnutrition: A Meta-Analysis of Demographic and Health Surveys Data. *International Journal of Environmental Research*

*name of corresponding author

and Public Health, *22*(3), 1–15. https://doi.org/10.3390/ijerph22030449

Raschka, S. (2020). *Model Evaluation 5 : Performance Metrics*.

Ratnasari, R., Wahidin, A. J., & Andika, T. H. (2024). Deteksi Dini Stunting Pada Anak Berdasarkan Indikator Antropometri dengan Menggunakan Algoritma Machine Learning. *Jurnal Algoritma*, *21*(2), 378–387. https://doi.org/10.33364/algoritma/v.21-2.2122

Sathyanarayanan, S., & Tantri, B. R. (2024). *Confusion Matrix-Based Performance Evaluation Metrics*. *27*(4).

Shen, H., Zhao, H., & Jiang, Y. (2023). Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea. *Children*, *10*(10). https://doi.org/10.3390/children10101638

Shirin Sara, S., Salauddin Khan, M., & Talukder, A. (2024). Prediction of Child Stunting with Machine Learning Algorithms: A Cross-Country Study of Bangladesh, India, and Nepal. *Midwifery.Iocspublisher.Org*. https://www.researchsquare.com/article/rs-4696630/latest

Sinaga, M., Fujiati, F., & Halawa, D. (2025). Designing a Stunting Prediction Model Using Machine Learning to Support SDGs Achievement in Indonesia. *Sinkron*, *9*(4), 2070–2079. https://doi.org/10.33395/sinkron.v9i4.15296

Sutarmi, S., Warijan, W., Indrayana, T., B, D. P. P., & Gunawan, I. (2023). Machine Learning Model For Stunting Prediction. *Jurnal Health Sains*, *4*(9), 10–23. https://doi.org/10.46799/jhs.v4i9.1073

Takahashi, K., Yamamoto, K., Kuchiba, A., & Koyama, T. (2022). *Confidence interval for micro-averaged F1 and macro-averaged F1 scores*. *52*(5), 4961–4972. https://doi.org/10.1007/s10489-021-02635-5.Confidence

Tamanna, T., Mahmud, S., Salma, N., Hossain, M. M., & Karim, M. R. (2025). Identifying determinants of malnutrition in under-five children in Bangladesh: insights from the BDHS-2022 cross-sectional study. *Scientific Reports*, *15*(1), 1–18. https://doi.org/10.1038/s41598-025-99288-y

UNICEF, & others. (2016). Levels and Trends in Child Malnutrition: UNICEF. *WHO/World Bank Group Joint Child Malnutrition Estimates*, 1–24.

Wicaksono, F., & Harsanti, T. (2020). Determinants of stunted children in Indonesia: A multilevelanalysis at the individual, household, and community levels. *Kesmas*, *15*(1), 48–53. https://doi.org/10.21109/kesmas.v15i1.2771

Zemariam, A. B., Abate, B. B., Alamaw, A. W., Lake, E. S., Yilak, G., Ayele, M., Tilahun, B. D., & Ngusie, H. S. (2025). Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine learning algorithms. *PLoS ONE*, *20*(1 January), 1–27. https://doi.org/10.1371/journal.pone.0316452

*name of corresponding author