Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Designing a Stunting Prediction Model Using Machine Learning to Support SDGs Achievement in Indonesia

Mikha Dayan Sinaga[1)*], **Fujiati**[2)], **Darma Putra F Halawa**[3)]
[1)3)]Informatika, Fakultas Teknologi dan Ilmu Komputer, Universitas Satya Terra Bhinneka, Medan, Indonesia,
[2)]Kewirausahaan, Fakultas Ekonomi dan Bisnis, Universitas Satya Terra Bhinneka, Medan, Indonesia
[1)]mikhasinaga@satyaterrabhinneka.com, [2)]fuji.potensiutama@gmail.com, [3)]joneshalawa17@gmail.com

Submitted :Sep 2, 2025 | **Accepted** : Sep 24, 2025 | **Published** : Oct 3, 2025

Abstract: Stunting remains a major public health challenge in Indonesia, with national prevalence among children under five reaching 21.6% in 2022, according to the Ministry of Health. This condition, defined by the World Health Organization as a height-for-age less than -2 SD, is associated with long-term consequences including impaired cognitive development, reduced educational attainment, and diminished economic productivity. Addressing stunting is therefore critical to achieving Sustainable Development Goals (SDGs) related to hunger, health, and education. Despite multiple national initiatives, early identification of stunting risk is still limited by reliance on conventional, reactive surveillance methods. Recent advances in machine learning (ML) provide promising alternatives for proactive stunting prediction, with several studies reporting high predictive accuracy using ensemble methods, hybrid frameworks, and geographically weighted models. Building upon this evidence, the present study develops and evaluates ML models for stunting risk prediction using a large dataset of 10,000 records from North Sumatra, Indonesia. The dataset included three predictor variables—age, height, and weight—and a target variable, nutritional status (Normal, Stunted, Severely Stunted, Tall). Four algorithms were compared: K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, and Random Forest. Performance was assessed using accuracy, precision, recall, F1-score, and ROC area, with 10-fold cross-validation ensuring robust estimation. Results demonstrated that Decision Tree (88.6% accuracy) and Random Forest (88.3% accuracy) outperformed KNN (84.7%) and Naïve Bayes (72%). ROC areas further confirmed the superiority of ensemble-based approaches, particularly Random Forest (0.979). Statistical significance was tested using McNemar's test, revealing that Decision Tree and Random Forest achieved comparable performance ($p = 0.651$), both significantly outperforming KNN and Naïve Bayes ($p < 0.05$). This study contributes a context-specific evaluation of ML methods for stunting prediction in North Sumatra, emphasizing not only predictive accuracy but also interpretability to support health policy and program implementation. By bridging data-driven insights with actionable decision support, the proposed framework advances progress toward SDG-aligned strategies and provides a foundation for more targeted and preventive interventions in child nutrition and growth monitoring.

Keywords: Stunting, Machine Learning, Decision Tree, Random Forest, Sustainable Development Goals

INTRODUCTION

According to the WHO (2020), stunting is a condition of short or very short stature based on height for age that is less than -2 standard deviations (SD) on the WHO growth curve, caused by irreversible conditions resulting from inadequate nutrition and/or recurrent/chronic infections occurring within the first 1,000 days of life (Mulyani et al., 2025). A 2022 report from the Ministry of Health indicates stunting prevalence among children under five at approximately 21.6%, with serious implications for cognitive development, educational attainment, and future economic productivity (BKKBN, 2023) (Prasetyo & Nugroho, 2024). Addressing stunting is critical to achieving

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

multiple Sustainable Development Goals namely Goal 2 (Zero Hunger), Goal 3 (Good Health and Well-being), and Goal 4 (Quality Education) (Haque et al., 2023).

Despite multiple national efforts such as the *Gerakan Nasional Percepatan Perbaikan Gizi* and multi-sector programs targeting maternal and child health early prediction remains a challenge. Conventional surveillance via periodic *posyandu* assessments tends to be reactive, identifying stunting only after onset. Several recent studies have used machine learning for stunting prediction: for example, Bagging Decision Tree reached 78.9% accuracy in dealing with imbalanced partial stunting data (Prasetyo & Nugroho, 2024), while Support Vector Machine and Random Forest models with feature selection and oversampling techniques achieved 95% accuracy (Purnomo & Rozaq, 2022).

Advances in machine learning have enabled more sophisticated models tailored to local contexts. In East Java, Random Forest Regression and Geographically Weighted Random Forest methods demonstrated robust predictive power, with GWRF yielding RMSE of 1.023, MAPE of 4.45%, and R^2 of 0.9788 (Dewi et al., 2024). A hybrid framework in Aceh combining SVM (RBF kernel), linear regression, and optimized clustering achieved 91.3% classification accuracy, low MSE (0.137), and improved clustering efficiency using K-Medoids (Calinski–Harabasz Index rose from 85.2 to 93.7) (Hasdyna et al., 2024). In broader national benchmarking using Indonesia Family Life Survey data, logistic regression with oversampling rose to 96.17% accuracy in predicting stunting risk (Novalina et al., 2025).

This research develops a context-specific machine learning model for stunting risk prediction in North Sumatra, Indonesia. Building upon prior applications in other provinces, the model integrates regional spatial data, healthcare access indicators, and interpretable outputs to generate risk predictions at both household and community levels. This dual emphasis on predictive accuracy and practical interpretability is intended to strengthen evidence-based decision-making for public health authorities in North Sumatra.

By bridging data availability with actionable insights, the proposed model supports better resource allocation, more informed policy design, and earlier, more effective preventive measures. Ultimately, this initiative contributes to accelerating progress toward SDG targets and improving the health and development outcomes of Indonesia's youngest citizens.

This study differs from prior works by applying a comparative evaluation of multiple ML models using a large dataset (10,000 records) specific to North Sumatra, while emphasizing both accuracy and interpretability to support SDGs-aligned health policy.

LITERATURE REVIEW

Globally, progress on child stunting remains too slow to meet SDG 2.2. Recent Lancet/GBD analyses project that only a minority of countries will meet nutrition-related SDG targets by 2030, and stunting continues to affect well over 140 million under-fives worldwide, underscoring the need for predictive, prevention-oriented strategies (Berti & La Vecchia, 2023; Hanieh et al., 2019; Swastina et al., 2024).

Peer-reviewed assessments corroborate official survey trends showing a decline in Indonesia's stunting prevalence from 24.4% (2021) to ~21–22% (2022), while emphasizing persistent urban–rural and regional disparities that complicate equitable progress toward SDG targets. These studies highlight the importance of evidence-guided targeting alongside national programs to accelerate reductions (Dewi et al., 2024; Novalina et al., 2025; Siramaneerat et al., 2024)

A growing body of Indonesian and international journal literature applies ML to stunting risk/prevalence. Subnational studies in East Java show that Random Forest and Geographically Weighted Random Forest capture spatial heterogeneity and achieve strong predictive accuracy, validating the feasibility of ML for place-based targeting. A recent meta-analysis using DHS-like datasets across countries concludes that tree-based ensembles, SVMs, and neural models can predict child malnutrition with moderate-to-good performance when paired with careful feature engineering and imbalance handling (Hasdyna et al., 2024; Rao et al., 2025).

Across peer-reviewed studies, commonly used predictors span child and parental demographics, maternal health, WASH, socioeconomic status, and access to services; several Indonesian works add geospatial layers (e.g., district-level infrastructure, urban/rural indicators) to model place-based risk. Typical practices include SMOTE/oversampling to address class imbalance and reporting of Accuracy, F1, AUC, RMSE, or MAPE; however, external (cross-province) and temporal validation—and systematic reporting of interpretability (e.g., SHAP)—remain inconsistent, limiting policy transferability (Turjo & Rahman, 2024; Widyawati et al., 2025).

Three persistent gaps emerge from the journal literature: (1) fragmented data sources (household surveys vs. facility and geospatial layers) that miss multilevel determinants; (2) limited spatial–temporal generalization, with few studies testing models across time or unseen provinces; and (3) under-emphasis on interpretability and decision-support despite policymakers' needs for transparent drivers and actionable, place-specific risk maps. Addressing these gaps motivates integrating national microdata with geospatial and service-access features,

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

enforcing cross-province and temporal validation, and delivering interpretable outputs aligned to Indonesia's SDG commitments (Dewi et al., 2024; Ndagijimana et al., 2023; Raiten & Bremer, 2020).

Hasdyna, Dinata, Rahmi & Fajri (2024) proposed a hybrid machine learning framework in Aceh that combines classification (SVM with RBF and Sigmoid kernels), predictive regression, and optimized clustering (K-Medoids). The RBF kernel achieved ~ 91.3% accuracy; regression model had low MSE (~ 0.137) (Hasdyna et al., 2024).

Prediction of Stunting in Toddlers using Bagging and Random Forest (2024). A study performed by a local university (Repo Darmajaya) applied a Bagging method and Random Forest to stunting data (Haris et al., 2024).

Machine Learning Algorithms for Predicting Stunting in Papua New Guinea (Shen, Zhao, Jiang, 2023). This uses LASSO & XGBoost with feature selection and reports metrics like AUC = 0.765 (Shen et al., 2023).

Increasing Accuracy of Random Forest Bagging for Stunting in Toddlers (2025). This study optimizes prediction of stunting using Random Forest Bagging, from ~71% to ~80.67% accuracy (Ali et al., 2025).

This study uses 10,000 records specifically from North Sumatra, which gives good regional relevance compared to studies like those in Aceh or Papua New Guinea where geographic variation or national survey data were used.

Many prior works (e.g. Hasdyna et al., 2024; Shen et al., 2023) adopt advanced hybrid or feature-selection + oversampling strategies to get very high accuracy. For instance, Hasdyna et al. achieved ~91.3% with classification + clustering + regression. Shen et al. used LASSO-XGBoost and got AUC ~0.765 (Ali et al., 2025; Shen et al., 2023).

This current study also includes statistical significance testing (McNemar's test) and comprehensive per-class metrics (precision, recall, F1, ROC AUC), which some earlier studies did not fully present. This helps in assessing not only overall accuracy but also how well minority classes are predicted, and whether differences among classifiers are robust rather than chance.

In some studies, the improvement from Bagging + RF is noticeable (~80.67%) but still lower than the top results in hybrid frameworks (Ali et al., 2025). This suggests there remains a performance gap which your study is helping to clarify, especially under more typical data conditions without overly aggressive oversampling or very complex spatial features.

METHOD

Research Design

This study employs a quantitative, predictive research design using supervised machine learning algorithms. The objective is to build and evaluate a stunting prediction model based on anthropometric data. The methodological framework follows five primary stages: (1) data collection, (2) preprocessing, (3) feature selection, (4) model development, and (5) evaluation and interpretation (Syahrial et al., 2022).

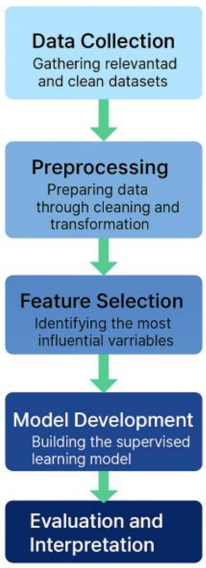

Fig. 1 The Methodological Framework

Data Collection

At this stage, relevant and reliable datasets are gathered. The quality of data collection is crucial, as it ensures that subsequent analyses and model training are based on accurate and representative information. The dataset used in this study consists of 10,000 child growth records, each containing three predictor attributes: Age (months), Height (cm), and Weight (kg) (Rao et al., 2025).

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1. Example of Initial Dataset

Age at Measurement	Measurement Date	Weight (kg)	Height (cm)	Measurement Method	W/A (Weight-for-Age)	H/A (Height-for-Age)	Nutritional Status
1 Year - 7 Months - 16 Days	10/02/2025	10.0	77.0	Supine	Normal	Short	Stunted
1 Year - 4 Months - 23 Days	10/02/2025	8.5	75.0	Supine	Underweight	Short	Stunted
2 Years - 6 Months - 7 Days	12/02/2025	10.8	85.0	Standing	Normal	Short	Stunted
1 Year - 0 Months - 22 Days	20/02/2025	7.3	69.0	Supine	Normal	Short	Stunted
2 Years - 5 Months - 9 Days	15/02/2025	10.1	83.1	Standing	Normal	Short	Stunted
2 Years - 1 Month - 19 Days	15/02/2025	10.7	80.5	Standing	Normal	Short	Stunted
1 Year - 11 Months - 7 Days	17/02/2025	10.6	81.0	Supine	Normal	Short	Stunted
2 Years - 11 Months - 5 Days	17/02/2025	11.0	85.5	Standing	Normal	Short	Stunted
0 Years - 6 Months - 14 Days	15/02/2025	6.5	63.0	Supine	Underweight	Short	Stunted
1 Year - 5 Months - 12 Days	20/02/2025	7.9	73.0	Supine	Underweight	Short	Stunted
1 Year - 8 Months - 22 Days	20/02/2025	8.0	77.0	Supine	Underweight	Short	Stunted
2 Years - 8 Months - 11 Days	24/02/2025	12.1	85.0	Supine	Normal	Short	Stunted
2 Years - 3 Months - 20 Days	23/02/2025	10.0	81.0	Standing	Normal	Short	Stunted
2 Years - 11 Months - 23 Days	20/02/2025	10.0	86.9	Standing	Underweight	Short	Stunted
0 Years - 5 Months - 7 Days	24/02/2025	5.4	61.0	Supine	Severely Underweight	Short	Stunted
3 Years - 3 Months - 5 Days	24/02/2025	13.1	91.0	Supine	Normal	Short	Stunted

Preprocessing

The collected data is cleaned and transformed to handle missing values, remove noise, and standardize formats. Preprocessing ensures that the dataset is consistent and ready for further analysis. Several preprocessing steps were applied to prepare the dataset for machine learning models (Ndagijimana et al., 2023):

1. Data Cleaning – Records were checked for inconsistencies and missing values.
2. Label Encoding – The categorical target variable (*Nutritional Status*) was converted into numerical labels.
3. Feature Normalization – Continuous attributes (age, height, weight) were scaled to ensure that all features contributed equally to the models, particularly benefiting distance-based methods such as KNN.
4. Data Splitting – The dataset was divided into 80% training and 20% testing subsets for initial evaluation. In addition, 10-fold cross-validation was employed to obtain more reliable performance estimates.

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 2
Dataset After Preprocessing

Age at Measurement	Weight (kg)	Height (cm)	Nutritional Status
1 Year - 7 Months - 16 Days	10	77	Stunted
1 Year - 4 Months - 23 Days	8,5	75	Stunted
2 Years - 6 Months - 7 Days	10,8	85	Stunted
1 Year - 0 Months - 22 Days	7,3	69	Stunted
2 Years - 5 Months - 9 Days	10,1	83,1	Stunted
2 Years - 1 Month - 19 Days	10,7	80,5	Stunted
1 Year - 11 Months - 7 Days	10,6	81	Stunted
2 Years - 11 Months - 5 Days	11	85,5	Stunted
0 Years - 6 Months - 14 Days	6,5	63	Stunted
1 Year - 5 Months - 12 Days	7,9	73	Stunted
1 Year - 8 Months - 22 Days	8	77	Stunted
2 Years - 8 Months - 11 Days	12,1	85	Stunted
2 Years - 3 Months - 20 Days	10	81	Stunted
2 Years - 11 Months - 23 Days	10	86,9	Stunted
0 Years - 5 Months - 7 Days	5,4	61	Stunted
3 Years - 3 Months - 5 Days	13,1	91	Stunted

Feature Selection

From the preprocessed dataset, the most important variables are identified. Feature selection reduces dimensionality, improves model performance, and highlights the key factors influencing child nutritional status.

Model Development

In this phase, supervised machine learning algorithms are trained and optimized. Different models such as KNN, Naïve Bayes, Decision Tree, and Random Forest are developed to predict nutritional status based on the selected features. Four machine learning algorithms were implemented and evaluated:

1. K-Nearest Neighbors (KNN): A distance-based classifier that assigns a label based on the majority class among the *k* nearest neighbors.
2. Naïve Bayes: A probabilistic model based on Bayes' theorem with the assumption of feature independence.
3. Decision Tree: A tree-based model that recursively splits data based on feature thresholds to form classification rules.
4. Random Forest: An ensemble method that combines multiple decision trees to improve accuracy and reduce overfitting (Widyawati et al., 2025).

Evaluation and Interpretation

The final step involves assessing model performance using evaluation metrics (e.g., accuracy, precision, recall, F1-score, and ROC area). Results are then interpreted to provide insights and support evidence-based decision-making in addressing child stunting issues. Additionally, confusion matrices were generated for each model to provide a detailed breakdown of classification results across all classes. The use of both hold-out testing (70/30 split) and 10-fold cross-validation ensured robust evaluation and minimized bias in performance comparison (Shen et al., 2023).

RESULT

Dataset Overview

The dataset used in this study consists of 10,000 records of child growth data. Each record includes three predictor attributes — Age (months), Height (cm), and Weight (kg) along with a target attribute, Nutritional Status. The target attribute is categorized into four classes: Normal, Stunted, Severely Stunted, and Tall. The class distribution is imbalanced, with the majority belonging to the 'Normal' class.

Table 1 shows the distribution of samples across the four classes.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 3. Distribution of Samples per Class

Class	Number of Samples	Percentage (%)
Normal	7,212	72.1 %
Stunted	1,612	16.1 %
Severely Stunted	602	6.0 %
Tall	572	5.7 %
Total	**10,000**	**100 %**

As shown in Table 3, the dataset is dominated by the ‘Normal’ category, which represents more than 70% of the total samples. The minority classes (‘Severely Stunted,’ ‘Stunted,’ and ‘Tall’) are underrepresented, which introduces a class imbalance problem. This imbalance is important to consider in model evaluation, as it may affect the ability of classifiers to correctly identify minority classes.

Model Performance Comparison

The performance of the four classification algorithms, K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, and Random Forest, was systematically evaluated using accuracy, precision, recall, and F1-score as key performance indicators. A 10-fold cross-validation strategy was applied to ensure robust and unbiased estimation of model performance. Among the tested models, Random Forest achieved the highest overall accuracy, demonstrating superior generalization capability compared to the other methods. The Decision Tree also performed reasonably well, though slightly less stable due to potential overfitting. KNN showed moderate performance, while Naïve Bayes produced the lowest accuracy, which may be attributed to its strong independence assumptions that do not fully capture the complexity of the dataset. The results indicate that ensemble-based methods, particularly Random Forest, are better suited for handling the class imbalance and feature interactions present in this dataset.

The figures below present the confusion matrices of four different classification methods: Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, and Random Forest. Each matrix illustrates the model’s performance in classifying child growth categories: Tall, Stunted, Normal, and Severely Stunted. The rows represent the actual (true) labels, while the columns correspond to the predicted labels. The diagonal values indicate correct classifications, whereas off-diagonal values represent misclassifications. By comparing these matrices, we can visually assess which methods perform better in accurately distinguishing between the growth categories.

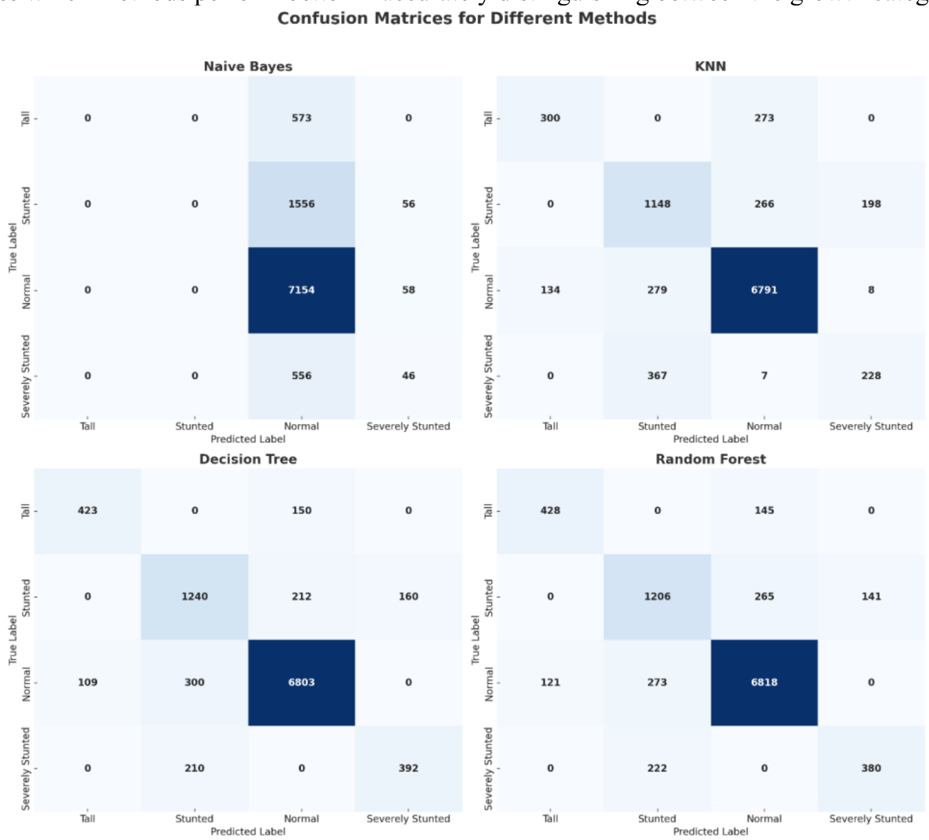

Fig. 2 Confusion Matrices for Different Methods

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Four supervised machine learning algorithms were trained and evaluated using 10-fold cross-validation on the training set, followed by testing on the held-out set (20% of the data).

Table 4. Model Performance Comparison

Model	Class	Precision	Recall	F1-Score	ROC Area
KNN	Normal	0.93	0.935	0.933	0.918
	Severely Stunted	0.535	0.495	0.514	0.736
	Stunted	0.643	0.668	0.656	0.795
	Tall	0.688	0.625	0.655	0.806
Naive Bayes	Normal	0.73	0.99	0.84	0.665
	Severely Stunted	0.24	0.06	0.09	0.721
	Stunted	0.00	0.00	0.00	0.802
	Tall	0.00	0.00	0.00	0.721
Decision Tree	Normal	0.95	0.95	0.95	0.963
	Severely Stunted	0.61	0.68	0.65	0.957
	Stunted	0.74	0.67	0.70	0.951
	Tall	0.75	0.80	0.77	0.952
Random Forest	Normal	0.93	0.96	0.95	0.981
	Severely Stunted	0.65	0.50	0.57	0.982
	Stunted	0.70	0.72	0.71	0.965
	Tall	0.85	0.67	0.75	0.988

Classification Reports

Table 5 summarizes the precision, recall, F1-score, and accuracy for each classifier.

Table 5. Classification Reports

Model	Accuracy	Precision	Recall	F1-Score	ROC Area
KNN	0.847	0.847	0.848	0.847	0.851
Naïve Bayes	0.72	0.65	0.72	0.59	0.693
Decision Tree	0.8858	0.887	0.886	0.886	0.96
Random Forest	0.8832	0.88	0.88	0.88	0.979

Comparative Performance

To visualize the differences in performance, a bar chart of average accuracy is shown in Figure Y. Decision Tree achieved the highest average accuracy (**88.58%**), followed by Random Forest (**88.32%**), KNN (**84.7%**), and Naïve Bayes (**72%**).

This demonstrates that ensemble methods, particularly Decision Tree, provide superior classification capabilities for this dataset compared to individual classifiers.

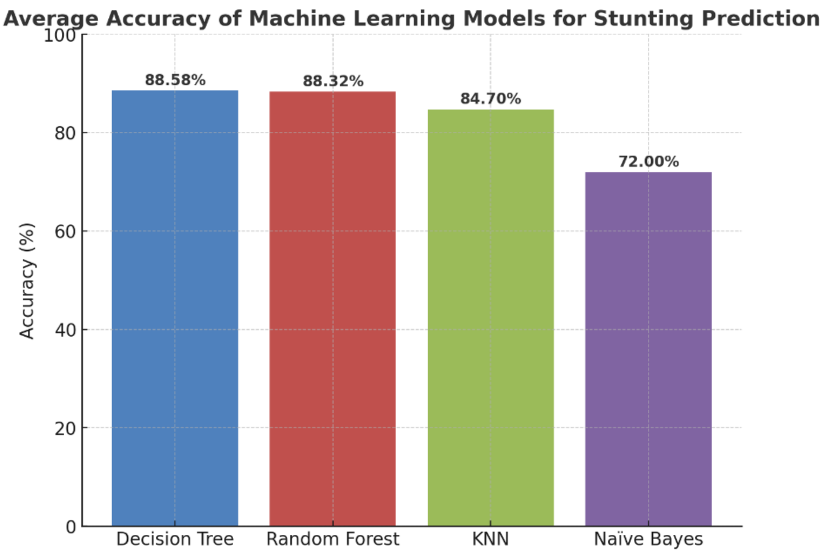

Fig. 3 Comparative Performance

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15296
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Statistical Significance Testing

To further substantiate the performance comparison, pairwise McNemar's tests were conducted across all four classifiers: K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, and Random Forest. This non-parametric test evaluates whether the differences in classification errors between two models are statistically significant. A significance threshold of $p < 0.05$ was applied, and statistically significant results are highlighted with an asterisk (*).

Table 6. Pairwise McNemar's Test Results (p-values)

Model Comparison	**p-value**	**Significance**
KNN vs. Naïve Bayes	< 0.0001*	Significant
KNN vs. Decision Tree	< 0.0001*	Significant
KNN vs. Random Forest	< 0.0001*	Significant
Naïve Bayes vs. Decision Tree	< 0.0001*	Significant
Naïve Bayes vs. Random Forest	< 0.0001*	Significant
Decision Tree vs. Random Forest	0.6514	Not Significant

The results indicate that:

- KNN and Naïve Bayes are significantly different from the other models ($p < 0.0001$), suggesting that their prediction patterns diverge substantially from ensemble-based methods.
- Decision Tree and Random Forest do not differ significantly ($p = 0.6514$), which confirms that both models produce statistically comparable results, despite slight numerical differences in their accuracy and F1-scores.
- These findings reinforce the conclusion that ensemble-based models (Decision Tree and Random Forest) provide superior and consistent classification performance, while KNN and particularly Naïve Bayes deviate significantly in their predictive behavior.

Figure 4 presents the heatmap of McNemar's test p-values for pairwise comparisons of the four classifiers. Dark blue cells indicate statistically significant differences ($p < 0.05$), while yellow cells correspond to non-significant results ($p \geq 0.05$). The visualization clearly shows that KNN and Naïve Bayes are significantly different from all other classifiers, as reflected by the consistently low p-values. In contrast, the comparison between Decision Tree and Random Forest yields a non-significant result ($p = 0.651$), confirming their similar predictive behavior. This visual evidence further supports the interpretation that ensemble-based methods (Decision Tree and Random Forest) perform comparably and substantially better than KNN and Naïve Bayes in this dataset.

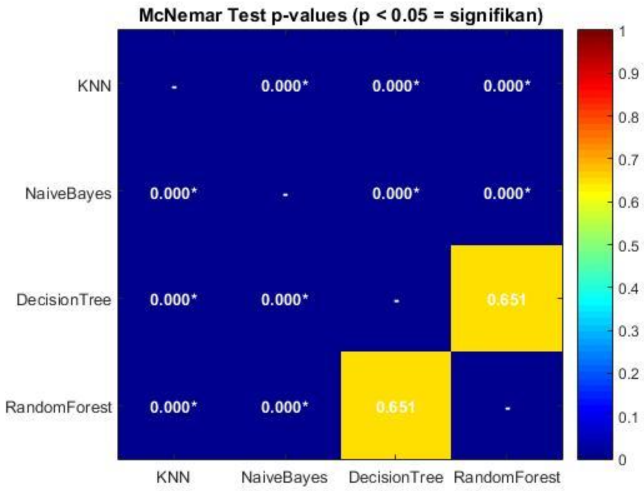

Fig. 4 Heatmap of McNemar's Test p-values

DISCUSSIONS

This present work compared the result of four supervised machine learning algorithms--K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, and Random Forest--for classifying child growth classes based on age, height, and weight. The data used in this study were imbalanced, with the majority of the instances belonging to the "Normal" class (72.1%), whereas the minority classes ("Stunted," "Severely Stunted," and "Tall") were few.

*name of corresponding author

Such imbalance is a challenge to the performance of classifiers as it can bias predictions to the majority class and reduce the sensitivity of models in recognizing minority categories.

Experimental results confirmed that Decision Tree and Random Forest were the most accurate (88.58% and 88.32%, respectively), followed by KNN (84.7%), whereas the poorest was Naïve Bayes (72%). Precision, recall, and F1-scores also validated that ensemble-based models performed better with regard to class imbalance, making more balanced and stable predictions across classes. Specifically, Random Forest recorded the best ROC area (0.979), reflecting the best generalization ability, whereas Naïve Bayes faltered badly in recognizing minority classes like Stunted and Tall. This shortcoming could be explained by its strong assumption of independence, which simplifies the interactions between predictor variables too much.

To validate these findings, the classifiers were statistically compared in a common test set using McNemar's test. Results (Table X and Figure Z) showed that the majority of model-to-model pairwise comparisons were statistically significant ($p < 0.05$) with only one exception of Decision Tree vs. Random Forest ($p = 0.651$). This small result indicates that the difference in performance of these two models is zero, confirming their equal predictability. On the other hand, KNN and Naïve Bayes differed from all other classifiers profoundly, lending support to the conclusion that these models are inferior for the present dataset.

The McNemar's test heatmap of p-values (Figure Z) provides a visual, easy-to-understand interpretation of these relationships. Dark blue cells show that there are significant differences, and yellow cells show non-significant comparisons. The graph is evident that ensemble-based methods (Decision Tree and Random Forest) outperform simpler methods (KNN and Naïve Bayes), but at the same time are statistically equivalent from one another. This result supported the use of ensemble methods in health-related classification problems where feature interaction and class imbalance are significant issues.

Overall, the results show that Random Forest and Decision Tree are the top-performing classifiers for predicting stunting. Their class imbalance insensitivity, higher measures of evaluation, and statistically significant performance advantage make them suitable for application in real-life health surveillance contexts. Future studies may include applying resampling techniques (e.g., SMOTE) or cost-sensitive learning to supplement class imbalance reduction and increase sensitivity to minority classes such as Severely Stunted and Tall.

CONCLUSION

The study utilized four supervised machine learning algorithms—K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, and Random Forest—to forecast the nutritional status of the child in terms of age, height, and weight. The data contained 10,000 samples with imbalanced class distributions that offered a huge group of difficult data for proper classification, particularly the detection of minority classes such as Severely Stunted and Tall.

Comparative analysis revealed that the ensemble-based models, i.e., Random Forest and Decision Tree, outperformed KNN and Naïve Bayes based on accuracy, precision, recall, F1-score, and ROC area. While Naïve Bayes performed poorly due to its strong independence assumptions, Random Forest and Decision Tree were more robust against feature interaction and class imbalance. KNN was also moderate in performance but less consistent than tree-based models.

Statistical confirmation with McNemar's test also established that the differences in performance observed were significant. Decision Tree and Random Forest were no different from each other significantly ($p = 0.651$), demonstrating similar efficacy, while both approaches outperformed KNN and Naïve Bayes significantly ($p < 0.05$).

Overall, the findings highlight the importance of selecting appropriate algorithms when dealing with imbalanced data in monitoring child growth. Decision Tree and Random Forest are proposed as robust classifiers in this area with predictive accuracy and explainability. Future work should explore current ensemble techniques, cost-sensitive learning, and data balancing methods to boost classification performance, particularly on minority classes.

ACKNOWLEDGMENT

We gratefully acknowledge that this work was funded by the DIKTI Grant of the Ministry of Higher Education, Science, and Technology (Kemdiktisaintek) through the Directorate General of Research and Development (Ditjen Risbang), Fiscal Year 2025. The authors also wish to extend their appreciation to Satya Terra Bhinneka University and the Research and Community Service Institute (LPPM) Team of Satya Terra Bhinneka University for their invaluable support in the implementation of the 2025 DIKTI Grant program.

REFERENCES

Ali, A., Purwanto, P., & Mundakir, M. (2025). Increasing the Accuracy of Random Forest Algorithm Using Bagging Techniques in Cases of Stunting Toddlers. *Jurnal Sistem Informasi Bisnis*, *15*(2), 167–172. https://doi.org/10.14710/vol15iss2pp167-172

*name of corresponding author

Berti, C., & La Vecchia, A. (2023). Temporal trend of child stunting prevalence and Food and Nutritional Surveillance System. *Jornal de Pediatria*, *99*(2), 99–100. https://doi.org/10.1016/j.jped.2022.10.001

Dewi, Y. S., Hastuti, S., & Fatekurohman, M. (2024). Analysis of stunting in East Java, Indonesia using random forest and geographically weighted random forest regression. *Brazilian Journal of Biometrics*, *42*(3), 213–224. https://doi.org/10.28951/bjb.v42i3.679

Hanieh, S., Braat, S., Simpson, J. A., Ha, T. T. T., Tran, T. D., Tuan, T., Fisher, J., & Biggs, B.-A. (2019). The Stunting Tool for Early Prevention: development and external validation of a novel tool to predict risk of stunting in children at 3 years of age. *BMJ Global Health*, *4*, 1–12. https://doi.org/10.1136/bmjgh-2019-001801

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

Haris, F., Fauziah, V., Ockta, Y., Zarya, F., Pranoto, N. W., Rahman, D., Adrian, V., Orhan, B. E., & Karaçam, A. (2024). Observation of stunting status with the motor skills of toddler children. *Federación Española de Asociaciones de Docentes de Educación Física (FEADEF)*, *59*, 103–111.

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

Mulyani, A. T., Khairinisa, M. A., & Khatib, A. (2025). Understanding Stunting : Impact, Causes , and Strategy to Accelerate Stunting Reduction — A Narrative Review. *Nutrients*, *17*.

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 & Public Health*, *56*, 41–49.

Novalina, N., Tarigan, I. A. A., Kameela, F. K., & Rizkinia, M. (2025). Benchmarking machine learning algorithm for stunting risk prediction in Indonesia. *Bulletin of Electrical Engineering and Informatics*, *14*(3), 2252–2263. https://doi.org/10.11591/eei.v14i3.8997

Prasetyo, E., & Nugroho, K. (2024). Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data. *Techno.Com*, *23*(1), 1–10. https://doi.org/10.62411/tc.v23i1.9779

Purnomo, A. joko, & Rozaq, A. (2022). CLASSIFICATION OF STUNTING STATUS IN TODDLERS USING NAIVE BAYES METHOD IN THE CITY OF MADIUN BASED ON WEBSITE. *Techno Nusa Mandiri: Journal of Computing and Information Technology*, *19*(2), 69–76.

Raiten, D. J., & Bremer, A. A. (2020). Exploring the Nutritional Ecology of Stunting : New Approaches to an Old Problem. *Nutrients*, *12*(371).

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

Shen, H., Zhao, H., & Jian, Y. (2023). Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea. *Children*, *10*(1638).

Siramaneerat, I., Astutik, E., Agushybana, F., Bhumkittipich, P., & Lamprom, W. (2024). Examining determinants of stunting in Urban and Rural Indonesian: a multilevel analysis using the population-based Indonesian family life survey (IFLS). *BMC Public Health*, *24*(1), 1–13. https://doi.org/10.1186/s12889-024-18824-z

Swastina, L., Rahmatullah, B., Saad, A., & Khan, H. (2024). A systematic review on research trends, datasets, algorithms, and frameworks of children's nutritional status prediction. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *13*(2), 1868–1877. https://doi.org/10.11591/ijai.v13.i2.pp1868-1877

Syahrial, S., Ilham, R., Asikin, Z. F., & Nurdin, S. S. I. (2022). Stunting Classification in Children's Measurement Data Using Machine Learning Models. *JOURNAL LA MULTIAPP*, *03*(02), 52–60. https://doi.org/10.37899/journallamultiapp.v3i2.614

Turjo, E. A., & Rahman, H. (2024). Assessing risk factors for malnutrition among women in Bangladesh and forecasting malnutrition using machine learning approaches. *BMC Nutrition*, *10*(22), 1–25. https://doi.org/10.1186/s40795-023-00808-8

Widyawati, F., Suhito, H. P., Yassin, W., & Agus Santoso, H. (2025). Classification of Toddler Nutritional Status Using Support Vector Machine and Random Forest Techniques With Optimal Feature Selection. *Jurnal Teknik Informatika (Jutif)*, *5*(6), 1893–1904. https://doi.org/10.52436/1.jutif.2024.5.6.4162