Institut Riset dan Publikasi Indonesia (IRPI)
IJATIS: Indonesian Journal of Applied Technology and Innovation Science
Journal Homepage: https://journal.irpi.or.id/index.php/ijatis
Vol. 3 Iss. 1 February 2026, pp: 29-40
ISSN(P): 3032-7466 | ISSN(E): 3032-7474

Comparison of Machine Learning Algorithm Performance for Toddler Stunting Prediction

Sophia Anjani[1*], Nadirah[2], Nur Qistina Binti Mohamad Iskandar[3], Mujahid Zinky[4], Ramzy Hammad Atmanagara[5], Rayhan Syahbani[6], Muhammad Marzuq[7], Muhammad Anis Fitri[8], Nur Khalis[9], Fajar Abiyyu Khairullah[10], Fawwaz Zanuar Alfarizy[11], Ahmad Fahiq Zauqol Kalam[12], Aldi Setiawan[13]

[1,2]Department of Information System, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[3]Diploma in Public Administration, Universiti Teknologi MARA, Malaysia
[4,7,11]Department of Islamic Studies, Shaikh Ahmad Kaftaro University, Syria
[5]Department of Sharia and Arabic Language, Cham International Islamic Center, Syria
[6,8,9]Department of Ushuluddin, Shaikh Ahmad Kaftaro University, Suriah
[10]Department of Sharia and Arabic Language, Cham International Islamic Center, Syria
[12]Department of Ushuluddin, Al Fath Al Islamic University, Syria
[13]Department of Sharia, Al Fath Al Islamic University, Syria

E-Mail: [1]12350321230@students.uin-suska.ac.id, [2]12350321372@students.uin-suska.ac.id, [3]2023831776@student.uitm.edu.my, [4]mujahidzinkyaddimasyqie@gmail.com, [5]atmangararamzy@gmail.com, [6]rayhansyahbani2002@gmail.com, [7]muhammadmarzuq1313@gmail.com, [8]anisahmad281103@gmail.com, [9]dmanjookholis@gmail.com, [10]fajarastech123@gmail.com, [11]fawwazalfarizy82@gmail.com, [12]fahiqzauqol@gmail.com, [13]kuncilemari123321@gmail.com

Received Jan 03rd 2026; Revised Mar 14th 2026; Accepted Mar 17th 2026; Available Online Mar 18th 2025
Corresponding Author: Sophia Anjani

Abstract

Stunting is a chronic nutritional issue in toddlers that has long-term effects on children's physical growth and cognitive development. This study aims to compare the performance of four machine learning algorithms, namely Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP), and Logistic Regression (LR), in classifying the nutritional status of toddlers. The research stages included data preprocessing, data division into training and test sets, model training, and evaluation using accuracy, precision, recall, F1-Score, a confusion matrix, and Area Under the Curve (AUC). The evaluation results showed that Random Forest achieved the best performance, with an accuracy of 94%, as well as precision, recall, and F1-score values above 90%, and an AUC value close to 1.00 across all nutritional status classes. This was followed by the MLP algorithm in second place, with an accuracy of 93.29%. The main contribution of this study is the identification of a high-performing, stable model for large-scale stunting detection, providing a strong foundation for developing decision-support systems for early detection in the public health sector.

Keywords: Early Detection, Machine Learning, Nutritional Status Classification, Random Forest, Toddler Stunting

1. INTRODUCTION

Stunting is a growth and development disorder experienced by children due to long-term nutritional deficiencies and dietary imbalances, which then affect cognitive development and school and work performance as adults [1], [2]. According to WHO data, globally in 2022, there were an estimated 149 million children under the age of 5 who were stunted (too short for their age), and 45 million toddlers were estimated to be wasted (too thin for their height).

In the last 10 years, the prevalence of stunting in Indonesia reached 30.8% in 2018 [3], and decreased to 21.5% in 2023 [4]. This condition of chronic malnutrition not only reduces height according to WHO standards but also increases the risk of child mortality, infectious diseases, brain development disorders, and metabolic problems in later life [5]. This condition underscores the need to detect stunting early through technological approaches, such as machine learning models, so that nutritional and health interventions can be provided earlier and more effectively. This study aims to compare the performance of four machine learning algorithms, namely Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP), and Logistic Regression, in classifying the nutritional status of toddlers.

Previous case studies have used machine learning to predict stunting. For example, Hendy et al. (2025) found that gradient Random Forest achieved the highest predictive performance with an accuracy score exceeding 90% [6], and Misganaw et al. (2025) reported that Random Forest achieved the highest performance with an accuracy of 97.985%, precision of 97.96%, recall of 97.965%, F1 score of 97.945%, and ROC-AUC of 99.995% [7]. Another study by Sabilillah et al. (2024) conducted a similar comparative analysis using the Kaggle 121K dataset and several algorithms, including Logistic Regression, SVM with an RBF kernel, CNN, and MLP. They found that SVM-RBF and CNN achieved the highest accuracy of 98%, and MLP achieved 97% [8], [9].

Unlike previous studies, this study uses a large stunting dataset from Kaggle that has not been widely explored in the literature. Our approach directly compares SVM, Random Forest, MLP, and Logistic Regression in predicting stunting on a broader scale. By comparing models on the same dataset, we hope to determine which algorithm achieves the best accuracy and evaluation metrics for detecting stunting. Essentially, this study aims to evaluate the advantages and disadvantages of each model in stunting cases, thereby supporting researchers and nutrition practitioners in selecting the most effective stunting prediction model. This study utilises a large, recent dataset and compares SVM, Random Forest, MLP, and Logistic Regression for stunting in toddlers, a topic that has not been widely researched previously.

This study uses the Toddler Stunting Detection dataset from Kaggle, which comprises 121,000 data points on toddlers with attributes such as age, gender, height, and nutritional status. The implementation was carried out in Python on the Google Colab platform, following the machine learning system flow. Then, the performance of each model was measured using a confusion matrix to obtain TP, TN, FP, and FN values, and using derivative metrics such as accuracy, precision, recall, and AUC. The results of this comparison are the main focus of this study.

2. MATERIAL AND METHOD

Several previous studies show that machine learning algorithms such as Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Logistic Regression perform well in classification tasks within the health sector. Research by Misganaw (2025) indicates that Random Forest can achieve the highest accuracy in predicting children's stunting status. Furthermore, research by Sabilillah (2024) also conducted a comparison of classification algorithms and found that SVM and MLP provide good performance in detecting stunting. Therefore, this study utilizes these four algorithms to conduct a comparative analysis to determine the most effective model for classifying toddler nutritional status.

This study uses the secondary dataset "Toddler Stunting Detection (121K rows)" from Kaggle (Rendiputra). The data were cleaned and normalised, and categorical attributes were encoded as numerical values. The data were then split into a training set and a test set. The analysis was performed in Google Colab using Python, with four machine learning algorithms: Support Vector Machine (SVM), Random Forest, Multilayer Perceptron (MLP), and Logistic Regression. The four models were trained using the training data, then tested and evaluated using metrics such as the confusion matrix, accuracy, precision, and recall [10]. Research Methodology can be view Figure 1.

Figure 1 shows the overall representation of the Machine Learning-based system, which consists of several main stages as follows:

1. Data Pre-Processing: This stage focuses on data cleaning to improve data quality before being used in the model training process. This stage includes cleaning the data to ensure there are no missing values or duplicates.
2. Data Split: In this stage, the dataset is divided using the hold-out method with three schemes: 80:20, 70:30, and 60:40. This aims to analyze the effect of the proportion of training data on model performance.
3. Classification Process: This stage involves the implementation of four Machine Learning-based classification algorithms, including Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP), and Logistic Regression (LR).
4. Evaluation: In this stage, a performance assessment is conducted for each specified algorithm. Evaluation is carried out using several performance metrics, namely accuracy, precision, recall, and F1-Score, calculated based on the confusion matrix. Additionally, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) are used to assess the model's ability to distinguish between classes.
5. Comparative Analysis: The final stage is a comparative analysis to evaluate the performance of all classifiers, enabling the selection of the optimal prediction model. The results of this comparison will then be used to determine the optimal prediction model for supporting an early stunting detection system.

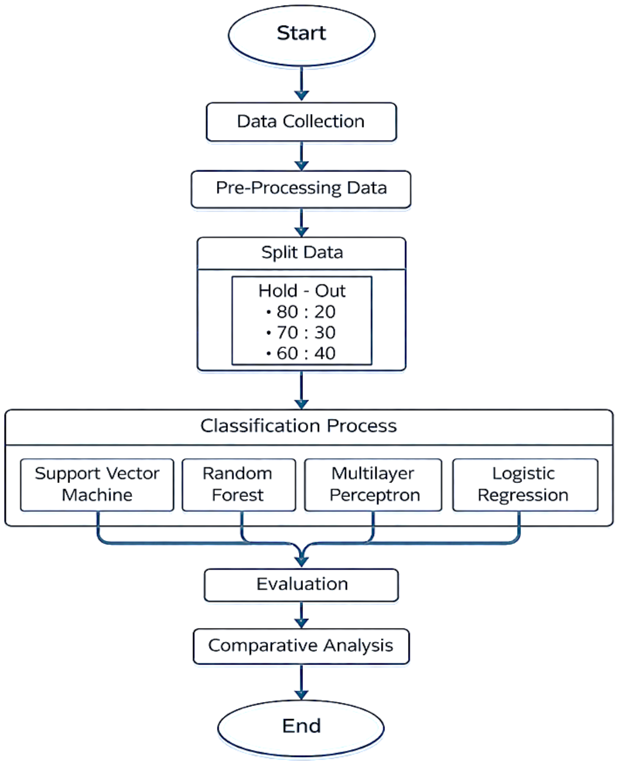

Figure 1. Research Methodology

2.1. Machine Learning

Machine Learning (ML) is a branch of artificial intelligence that enables computers to learn from data to make predictions or decisions without being explicitly programmed. In classification, ML models are trained with labelled data to recognise patterns that link input features to target classes [11], [12]. Machine Learning enables systems to recognise themselves and then predict outcomes. Of course, the effectiveness of ML algorithms depends heavily on feature selection and successful training [13]. Trained models can then be used to classify new data. The general process of ML classification includes data collection, pre-processing, model training, and evaluation using test data.

2.2. Support Vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm for classification and regression that seeks to separate data with a hyperplane of maximum distance between classes [14]. The SVM algorithm maps data to a high-dimensional feature space and constructs an optimal hyperplane that maximizes the margin between classes. In this way, only support vectors determine the model. SVM models are generally obtained by solving a quadratic programming problem, where the regularisation parameter C controls the trade-off between a large margin and classification errors on the training data. As a result, SVM produces a linear decision function (in feature space) that can be used to predict the class of new data [15]. In this case, the general SVM formula can be written as equation 1.

$$f(x) = \mathrm{sign}(w \cdot x + b) \tag{1}$$

Where f(x) is the prediction function, w is the normal vector to the hyperplane, x is the input feature vector, and b is the bias or intercept value [16].

2.3. Random Forest (RF)

Random Forest is an ensemble method that combines many random decision trees to improve prediction accuracy and stability [17]. Breiman (2001) explains that each tree in the forest is constructed by taking random samples (with replacement) from the training data and selecting random feature subsets for each node split. After all trees are built, the final prediction (classification) is obtained through majority voting of the trees. This approach reduces overfitting that often occurs in single trees and effectively reduces model variance. Breiman notes that the generalisation error in Random Forest depends on the strength of the individual trees and the correlation between trees with feature randomisation, the forest proves to be more robust to noise and provides high accuracy [18]. At each node in the tree, the algorithm determines which features to use for splitting based on criteria such as Gini Impurity or Information Gain. Gini Impurity is calculated using Formula, while Information Gain uses Formula [19], as shown in the formula 2.

$$\mathrm{Entropy(D)} = \sum_{l}^{k} = 1 \ \mathrm{P1 \log2 \ (P1)} \tag{2}$$

Where P1 describes the probability of each class in the dataset. Alternatively, Information Gain is calculated by comparing the entropy values before and after separation, as shown in formula 3.

$$\mathrm{IG\ (D,A) = Entrophy\ (D)\ -}\textstyle\sum_{v} \in \mathrm{Values(A)} \frac{|D|}{D} \mathrm{\ enthrophy\ (D_v)} \tag{3}$$

2.4. Multilayer Perceptron (MLP)

A Multilayer Perceptron is a feedforward artificial neural network consisting of an input layer, one or more hidden layers, and an output layer. Each unit (neuron) in the hidden and output layers calculates a linear combination of its inputs plus a bias, then applies a nonlinear activation function (usually differentiable). With this hidden layer, MLP can model complex nonlinear relationships between features and targets [20]. A characteristic feature of MLP is its unidirectional data flow: inputs flow from the input layer to the output layer without feedback loops. The MLP training process is generally carried out using a backpropagation algorithm with gradient descent, where the difference between the predicted output and the target is calculated and backpropagated to update the network weights, thereby reducing the error [21]. The formula for MLP can be written as an equation 4 [22].

$$\mathrm{Zj} = \sum_{i}^{n} = 1 \ \mathrm{Wji\ Xi} + \mathrm{bj} \tag{4}$$

2.5. Logistic Regression (LR)

Logistic regression is a statistical analysis technique used to model the relationship between a categorical outcome variable and one or more explanatory variables. This technique is generally used when the outcome variable (Y) has two categories, such as 0 and 1 (sick or healthy) or stunted or not stunted [23]. Logistic regression can produce the probability of a category occurring and make decisions based on an S-shaped logistic curve [24]. This model uses the following formula to determine the probability of belonging to a particular class. The probability of a class can be calculated using the sigmoid function, as in the following equation 5.

$$\mathrm{P(Y = 1|X)} = \frac{1}{1+3-z} \tag{5}$$

Where z is defined as a linear combination of input features using the equation 6 [15].

$$\mathrm{Z = W0 + w1x1 + w2x2 + \ldots . + wnxn} \tag{6}$$

2.6. Confusion Matrix

After the model is trained, its performance is evaluated using a confusion matrix, which shows the counts of correct predictions (True Positive, True Negative) and incorrect predictions (False Positive, False Negative). From this matrix, metrics such as accuracy (the ratio of correct predictions to the total data), precision (TP/(TP+FP)), recall/sensitivity (TP/(TP+FN)), and F1-score are calculated. The ROC (Receiver Operating Characteristic) curve and its area under the curve (AUC) are also used to assess the model's discriminatory ability at different decision thresholds. AUC is particularly useful when classes are imbalanced, as it evaluates the trade-off between True Positive Rate and False Positive Rate. By comparing these metrics, we can assess which model is the most accurate and reliable in detecting stunting [12].

2.7. Toddler Stunting

The case studied is the prediction of stunting in toddlers using a secondary dataset. Stunting is a condition characterised by growth retardation (below the age-standard) due to chronic malnutrition [9], [15]. The Kaggle dataset used includes 121,000 rows of toddler data with features such as age, gender, height, and nutritional status. The objective of this study is to classify toddlers into the categories "Severe stunting," "Stunting," "Normal," and "Tall" based on these attributes. The use of a large dataset is expected to improve the model's accuracy and robustness. As in previous studies, ML methods are expected to identify the key factors underlying stunting and to produce a predictive model that supports child health interventions.

3. RESULTS AND DISCUSSION

This section presents the results of experiments applying several classification algorithms to the Stunting Toddler Detection dataset. Model performance was evaluated using accuracy, precision, recall, and F1-Score metrics by applying three training and testing data division schemes, namely 80:20, 70:30, and 60:40, to analyse the effect of data ratio on model performance and stability. The evaluation results were then analysed to identify the algorithm with the best performance in predicting toddler nutritional status.

3.1. Data Collection

The dataset used in this study is the Stunting Toddler Detection dataset, which contains basic information on toddlers' characteristics and nutritional status. The data was obtained from verified sources and compiled in tabular format to support the analysis and classification modelling process. This dataset consists of 121,000 toddler data points with four main features and one target variable for classifying nutritional status.

The age and height variables play an important role in identifying children's growth patterns and detecting indications of stunting. Meanwhile, gender is used as a categorical variable to distinguish growth standards, as there are differences in anthropometric characteristics between boys and girls. Finally, nutritional status is the target variable that classifies toddlers' condition. The collected data was then checked to ensure there were no missing values, duplicates, or format inconsistencies. All numerical features were also evaluated to ensure unit and scale consistency. Information about each feature in the dataset is presented in Table 1.

Table 1. Initial Data

Column Name	Data Type	Description
Age (Months)	Numeric	Indicates the age of an infant or toddler in months (0–60 months). This variable is important for assessing growth phases and comparing a child's development with applicable age standards.
Gender	Categorical	Contains two categories, namely male and female. This variable is used to distinguish growth patterns because anthropometric standards differ between genders.
Height (cm)	Numeric	Records the height of toddlers in centimeters. It is a key indicator in linear growth analysis and is used to determine whether a child deviates from growth standards.
Nutritional status (Label)	Categorical	This is a target variable that classifies a child's nutritional status into four categories: severely stunted, stunted, normal, and tall, based on the WHO standard z-score. It is used in the process of predicting nutritional status.

3.2. Pre-Processing Data

The pre-processing stage involves data cleaning to ensure the quality of the Toddler Stunting Detection dataset is suitable for machine learning. This process includes missing values, noise, and outliers to ensure the data is consistent and ready for use in the next stage of analysis. The examination results show that all data in the dataset are complete, with no missing values. Dataset Pre-Processing can be seen in Table 2.

Table 2. Dataset Pre-Processing

Age (Months)	Gender	Height (cm)	Nutritional Status
0	Male	44.591973	Stunted
0	Male	56.705203	Height
0	Male	46.863358	Normal
0	Male	47.508026	Normal
0	Male	42.743494	Severely stunted

Categorical attributes, namely gender and nutritional status, are converted to numerical values using Label Encoding to enable machine learning algorithms to process the data. Numerical attributes such as age and height are normalised using StandardScaler to equalise the data scale and improve the stability and performance of the classification algorithm used.

3.3. Feature Selection

After preprocessing, feature selection was performed using a correlation-based method. This method was applied to analyse the relationship between features in the Stunting Toddler Detection dataset, which included four features and one target variable. This method measures the degree of linear relationship between features by calculating the absolute correlation coefficient. Features that are highly correlated with other features tend to provide similar information and can lead to multicollinearity issues in analyses or predictive models. The results of these calculations are visualised as a heatmap in Figure 2.

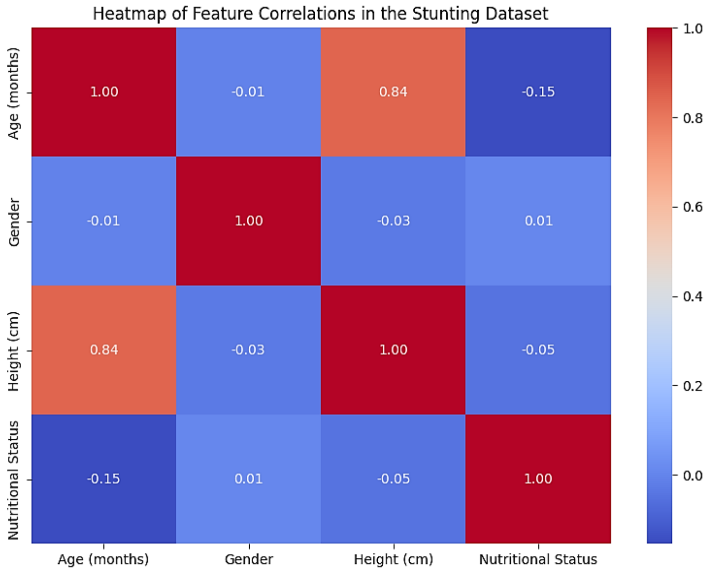

Figure 2. Heatmap of feature correlations in the stunting dataset

Figure 2 shows the relationship between features in the stunting dataset. Darker red indicates a significant positive correlation, while darker blue indicates a negative correlation or a very weak relationship. From the heatmap, it can be seen that the Age (months) and Height (cm) features have a fairly high correlation level of 0.84, indicating a strong linear relationship between the two. Conversely, the Gender feature shows a very low correlation with almost all other features, indicating that this variable does not exhibit a meaningful linear relationship in the dataset.

In feature selection methods based on correlation, features that exhibit very high correlation with each other can be removed to reduce information redundancy. On the other hand, features with low correlations, such as Gender, are better retained because they contain unique information that can support the model's prediction of Nutritional Status. This heatmap visually depicts the pattern of relationships among features that are important during the feature analysis and selection stage.

Based on the selected features, the next step is to evaluate the performance of several classification algorithms to determine which model is most effective in predicting Nutritional Status. Therefore, a comparison was made among four algorithms: Support Vector Machine (SVM), Random Forest, Multi-Layer Perceptron (MLP), and Logistic Regression. The following graph compares the performance of each algorithm across four main metrics: Accuracy, Precision, Recall, and F1-Score.

3.4. Support Vector Machine (SVM)

Table 3 presents the results of evaluating the Support Vector Machine algorithm's performance across various data division schemes. The test results show that variations in the training-to-test data ratio do not significantly affect model performance. Among the schemes tested, the 70:30 and 60:40 data divisions showed optimal performance, particularly in maintaining a balance between classification ability and model consistency. Overall, SVM is resilient to changes in data partitioning schemes, though its performance remains moderate in classifying toddlers' nutritional status.

Table 3. SVM Evaluation Results

Ratio	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
80:20:00	78	78	77	76
70:30:00	78	78	77	77
60:40:00	78	78	77	77

3.5. Random Forest

Table 4 presents the results of evaluating the Random Forest algorithm's performance across various data division schemes. The test results show that the Random Forest algorithm produces very consistent and stable performance across all tested data ratios. There were no significant differences in performance across the 80:20, 70:30, and 60:40 schemes, indicating that changes in the training data proportion did not significantly affect the model's ability to generalize. These findings show that Random Forest has good generalization capabilities and superior performance in predicting the nutritional status of toddlers.

Table 4. RF Evaluation Results

Ratio	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
80:20:00	94	91	91	91%
70:30:00	94	91%	91	91
60:40:00	94	91%	91	91

3.6. Multi-Layer Perceptron (MLP)

As shown in Table 5, the results of the Multi-Layer Perceptron algorithm evaluation on several data division schemes are presented. The test results show that the MLP algorithm produces relatively stable performance across all data ratios applied. Among the schemes tested, the 70:30 ratio showed the most optimal and consistent performance, while the 60:40 ratio produced relatively lower performance.

Table 5. MLP Evaluation Results

Ratio	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
80:20:00	93	90	90	90
70:30:00	93	89	91	90
60:40:00	93	89	90	90

3.7. Logistic Regression

Table 6 shows the results of evaluating the performance of the Logistic Regression (LR) algorithm on several data division schemes. The test results show that the LR algorithm's performance is consistent across all specified data ratios. Among the schemes tested, the 60:40 ratio showed the most optimal performance, while the 80:20 and 70:30 schemes produced slightly lower but still comparable performance. Overall, these findings indicate that Logistic Regression can provide reliable classification results and is not sensitive to variations in data splitting.

Table 6. Evaluation Results of LR

Ratio	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
80:20:00	78	78	77	77
70:30:00	78	78	78	77
60:40:00	79	79	78	77

3.8. Algorithm Performance Comparison Analysis

The performance of the four machine learning algorithms, namely Random Forest, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Logistic Regression, was conducted using three data division schemes, namely 60:40, 70:30, and 80:20. This analysis aims to assess the effect of the test data proportion on the performance of each algorithm and determine the most optimal data split scheme. The comparison was made using accuracy, precision, recall, and F-1 score metrics, which are displayed in graph form for each data split scheme. Comparison of algorithm performance using the 60:40 data split scheme can view Figure 3.

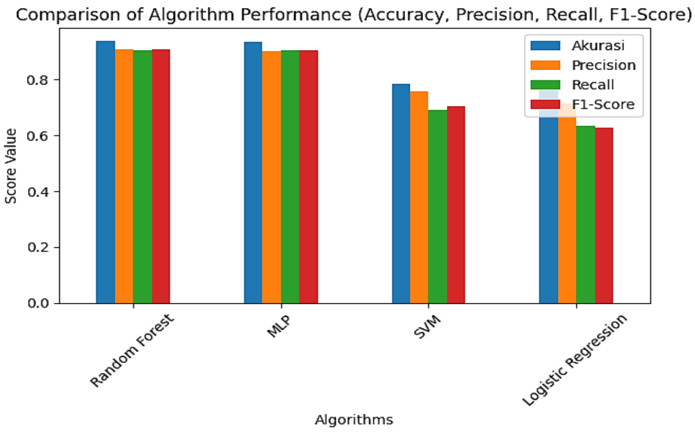

Figure 3. Comparison of algorithm performance using the 60:40 data split scheme

In Figure 3, the 60:40 data split scheme shows a downward trend in performance for several algorithms. This decline is particularly evident in the recall and F1-score metrics, indicating a reduction in the model's ability to correctly identify all classes due to the smaller training data set. Nevertheless, Random Forest and MLP still show relatively stable performance and remain superior to other algorithms. Conversely, SVM and Logistic Regression experience a more pronounced decline, particularly in recall, indicating that these two algorithms are more sensitive to reductions in the amount of training data. These results confirm that a limited amount of training data can affect a model's generalisation ability.

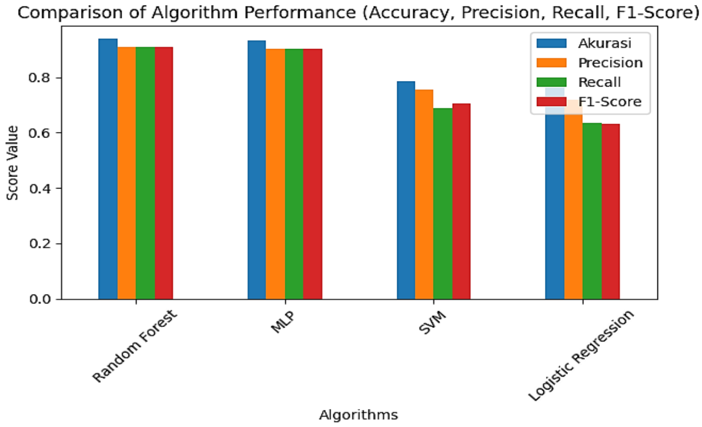

Figure 4. Comparison of algorithm performance using the 70:30 data split scheme

Based on Figure 4, the 70:30 data split scheme shows the most stable and consistent performance across all algorithms tested. The Random Forest algorithm achieved the highest scores across all evaluation metrics, with an accuracy of 0.94 and precision, recall, and F1-score values all above 0.90. These results indicate that Random Forest can achieve accurate, balanced classification across all nutritional status classes. The MLP algorithm also performed well and was relatively close to Random Forest, though there were still small differences in some metrics. Support Vector Machine (SVM) performed at a medium level, with a lower recall, indicating limitations in evenly recognising all classes. Meanwhile, Logistic Regression performed worst, especially in recall and F1-score, with values below 0.65, indicating the limitations of linear models in capturing complex data patterns. Overall, the 70:30 scheme provides an optimal balance between training and testing data, so it was selected as the best scheme.

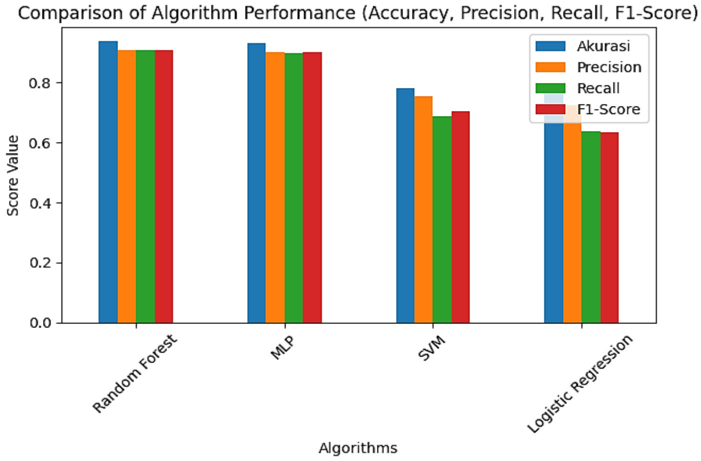

Figure 5. Comparison of algorithm performance using the 80:20 data split scheme

Based on Figure 5, the 80:20 data split scheme maintains high performance across algorithms with higher model complexity, especially Random Forest and MLP. The accuracy, precision, recall, and F1-score values for both algorithms remain high, though they do not show a significant improvement over the 70:30 scheme. On the other hand, SVM and Logistic Regression showed stagnant performance, with a significant gap in performance compared to Random Forest and MLP. This shows that although an increase in the amount of training data can help the model learn better, a test data proportion that is too small can reduce the representativeness of the performance evaluation.

Based on the evaluation results using three data distribution schemes, namely 70:30, 80:20, and 60:40, the 70:30 scheme showed the most stable and consistent performance across all algorithms tested. This scheme provides an optimal balance between the amount of training data and test data, thereby producing a more representative performance evaluation. Therefore, the 70:30 scheme was selected as the best scheme and used in the algorithm classification performance comparison stage.

These results indicate that differences in algorithm characteristics significantly affect the resulting classification performance. Algorithms such as Random Forests are better at generalisation because they can combine multiple decision models, making them more effective at capturing complex data patterns. Due to its limitations as a linear model, Logistic Regression tends to be unsuccessful on this dataset. Meanwhile, the moderate performance of SVM indicates that this algorithm still requires parameter tuning to improve classification performance across all classes, despite its ability to handle nonlinear data.

3.9. Confusion Matrix Analysis of the Best Model

Based on the model performance evaluation, Random Forest was identified as the model with the best and most stable performance across various data division scenarios. The advantages of this model are reflected in its ability to produce high accuracy and a good balance between precision, recall, and F1-score values for each nutritional status. Therefore, further analysis was conducted using a confusion matrix to better understand the model's ability to classify each toddler's nutritional status class.

As shown in Figure 6, the confusion matrix illustrates the Random Forest classifier's performance in classifying four classes of toddler nutritional status: Normal, Severely Stunted, Stunted, and High. The values on the main diagonal indicate correct predictions for each class, while those off the diagonal indicate classification errors. The model performed very well in the Normal and High classes, with a dominant rate of correct predictions.

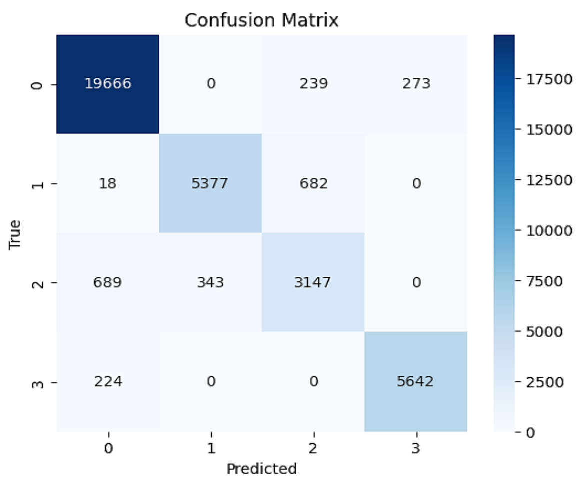

Figure 6. Confusion matrix of the classification results for the stunting dataset

In class 0 (Normal), the Random Forest model showed excellent performance, correctly classifying 19,666 data points, although there were a few errors in classes 2 and 3 due to their similar characteristics. Class 1 (Severely Stunted) was also detected well, with 5,377 correct predictions, but there were still errors in class 2, reflecting the overlap between severe and mild stunting characteristics. For class 2 (Stunted), the accuracy was relatively lower due to its nature as a transitional class, with some data misclassified into classes 0 and 1. Meanwhile, class 3 (Tall) was classified very accurately with 5,642 correct predictions, demonstrating the model's ability to consistently recognize high nutritional status.

3.10. Receiver Operating Characteristic (ROC) Curve Analysis of the Best Model

To further evaluate the discriminatory ability of Random Forest in distinguishing categories of toddler nutritional status, a Receiver Operating Characteristic (ROC) analysis was performed. Based on the test results in Figure 7, the Area Under the Curve (AUC) values for the Normal, Severely Stunted, and High classes each reached 1.00, while the Stunted class obtained an AUC value of 0.98. The ROC curve for all classes was close to the upper left corner of the graph, indicating a high True Positive Rate and a low False Positive Rate. The AUC value of 0.99 across all classes indicates that the model has very high accuracy and sensitivity in distinguishing between the positive and negative classes. These results confirm that Random Forest is highly effective and reliable at detecting stunting in toddlers across various nutritional status categories.

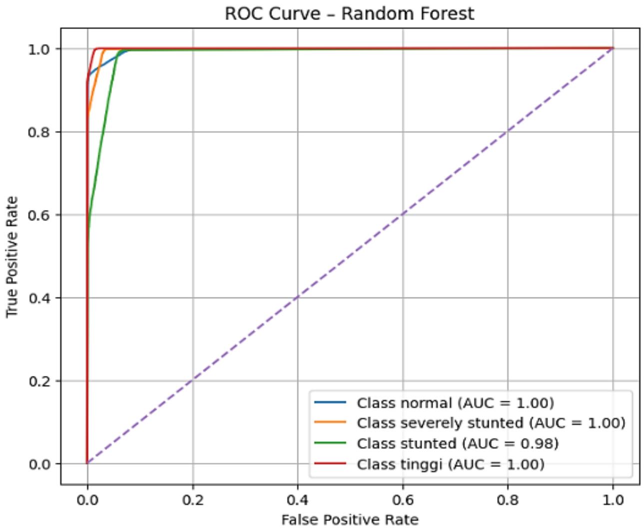

Figure 7. ROC Curve of the Random Forest algorithm for stunting status classification

3.11. Discussion

The comparison results show that the Random Forest (RF) algorithm achieved the best performance across all metrics (accuracy, precision, recall, F1-score), followed by the MLP algorithm, while SVM and LR achieved the lowest performance. These findings are consistent with the research of Misganaw (2025), which recorded an accuracy of 97.985% [7]. In a study by Nur'aini (2024), the Random Forest algorithm also produced the best results, with an accuracy of 99.95%, precision of 99.89%, recall of 99.94%, and F1-score of 99.88% [25]. Hendy also reported that the random forest algorithm achieved an accuracy exceeding 90% and an ROC-AUC value above 0.96 [26]. This consistency of results reinforces the position of RF as the most reliable algorithm for nutritional status classification across various data scales.

The superior performance of Random Forest in this study is heavily influenced by the characteristics of the dataset, which has complex non-linear relationships, such as the interaction between age, gender, and height relative to the z-score threshold. RF has advantages due to its ensemble mechanism, which reduces variance by combining many random decision trees, thus handling noise and data complexity effectively [27], This allows for more stable learning of non-linear patterns compared to linear models like Logistic Regression, which have limitations in capturing complex relationships between variables. Additionally, the bagging mechanism makes RF more robust against outliers that often appear in field data.in line with research by Sabilillah (2024) [3]

The findings of this study have significant implications. Random Forest has proven to be superior and has the potential to be implemented in automated stunting detection systems. Higher prediction accuracy allows for the identification of toddlers at risk of stunting at an earlier stage. The strength of this study lies in the use of a large dataset and comprehensive evaluation metrics.

4. CONCLUSION

This study compared the performance of four machine learning algorithms—Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Logistic Regression (LR)—in classifying toddler nutritional status to detect stunting. Evaluation results show that Random Forest provides the best performance, with the highest accuracy, precision, recall, and F1-score values across various data division schemes. Furthermore, confusion matrix and ROC-AUC analyses also indicate excellent classification ability with a low error rate. These findings suggest that Random Forest is the most effective algorithm for toddler stunting classification. Future research is suggested to optimize model parameters and expand the dataset, for example, by using national-scale health data or adding more diverse features to improve model accuracy.

REFERENCES

[1] A. Ayelign and T. Zerfu, "Household, dietary and healthcare factors predicting childhood stunting in Ethiopia," Heliyon, vol. 7, no. 4, Apr. 2021, doi: 10.1016/j.heliyon.2021.e06733.

[2] G. Nduwayezu, P. Zhao, P. Pilesjö, J. P. Bizimana, and A. Mansourian, "Multilevel small-area childhood stunting risk estimation: Insights from spatial ensemble learning, agro-ecological and environmentally remotely sensed indicators," Environmental and Sustainability Indicators, vol. 27, Sep. 2025, doi: 10.1016/j.indic.2025.100822.

[3] 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, Oct. 2024, doi: 10.33395/sinkron.v8i4.13967.

[4] M. As'an Hamid and R. Subhiyakto, "Performance Comparison of Random Forest, SVM, and XGBoost Algorithms with SMOTE for Stunting Prediction," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[5] I. A. Cahyani, P. I. Ashuri, and C. S. K. Aditya, "Stunting Disease Classification Using Multi-Layer Perceptron Algorithm with GridSearchCV," Sinkron, vol. 9, no. 1, pp. 392–401, Jan. 2024, doi: 10.33395/sinkron.v9i1.13245.

[6] A. Hendy et al., "Supervised machine learning for classification and prediction of stunting among under-five Egyptian children," BMC Pediatr, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12887-025-06138-x.

[7] M. K. Ayele, G. A. Baye, S. H. Yesuf, A. A. Engda, and E. T. Mitiku, "Predicting stunting status among under five children in ethiopia using ensemblemachine learning algorithms," Sci Rep, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-03206-1.

[8] 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, Oct. 2024, doi: 10.33395/sinkron.v8i4.13967.

[9] S. Syahrial, R. Ilham, Z. F. Asikin, and St. S. I. Nurdin, "Stunting Classification in Children's Measurement Data Using Machine Learning Models," Journal La Multiapp, vol. 3, no. 2, pp. 52–60, Mar. 2022, doi: 10.37899/journallamultiapp.v3i2.614.

[10] M. M. Islam et al., "Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh," International Journal of Cognitive Computing in Engineering, vol. 3, pp. 46–57, Jun. 2022, doi: 10.1016/j.ijcce.2022.02.002.

[11] D. Azriani, D. Agustian, Y. Zuhairini, I. N. Yulita, and M. Dhamayanti, "Prediction models for stunting at 2-years-old from Indonesian newborn population," BMC Pediatr, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12887-025-06096-4.

[12] A. B. Zemariam et al., "Employing supervised machine learning algorithms for classification and prediction of anemia among youth girls in Ethiopia," Sci Rep, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-60027-4.

[13] Ü. Ağbulut, A. E. Gürel, and Y. Biçen, "Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison," Renewable and Sustainable Energy Reviews, vol. 135, Jan. 2021, doi: 10.1016/j.rser.2020.110114.

[14] S. Yousefi, "Comparison of the performance of machine learning algorithms in predicting heart disease," Frontiers in Health Informatics, vol. 10, 2021, doi: 10.30699/fhi.v10i1.349.

[15] N. Amaya-Tejera, M. Gamarra, J. I. Vélez, and E. Zurek, "A distance-based kernel for classification via Support Vector Machines," Front Artif Intell, vol. 7, 2024, doi: 10.3389/frai.2024.1287875.

[16] A. Karatzoglou, D. Meyer, W. Wien, and K. Hornik, "Journal of Statistical Software Support Vector Machines in R," 2006. [Online]. Available: http://www.jstatsoft.org/

[17] M. Ghosh et al., "A comparative analysis of machine learning algorithms to predict liver disease," Intelligent Automation and Soft Computing, vol. 30, no. 3, pp. 917–928, 2021, doi: 10.32604/iasc.2021.017989.

[18] L. Breiman, "Random Forests," 2001.

[19] C. P. Silalahi and S. Lestari, "Advanced Machine Learning Implementation for Early Detection and Prediction of Alzheimer's Disease," Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM), vol. 8, no. 3, pp. 558–568, 2025, doi: 10.24014/ijaidm.v8i3.38004.

[20] S. Shamshirband et al., "Comparative analysis of hybrid models of firefly optimization algorithm with support vector machines and multilayer perceptron for predicting soil temperature at different depths," Engineering Applications of Computational Fluid Mechanics, vol. 14, no. 1, pp. 939–953, Jan. 2020, doi: 10.1080/19942060.2020.1788644.

[21] P. Sharma, N. Malik, N. Akhtar, and H. Rohilla, "FEEDFORWARD NEURAL NETWORK: A Review," International Journal of Advanced Research in Engineering and Applied Sciences, vol. 2, no. 10, 2013, [Online]. Available: www.garph.co.uk

[22] M. F. Ramadhan, S. Y. Irianto, and A. Farduwin, "Lithology Prediction Using Deep Learning Artificial Neural Network and Schlumberger Resistivity Inversion Data at Eastern Lampung," Indonesian Journal of Artificial Intelligence and Data Mining, vol. 8, no. 2, p. 427, Aug. 2025, doi: 10.24014/ijaidm.v8i2.37652.

[23] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," in IOP Conference Series: Materials Science and Engineering, IOP Publishing Ltd, Nov. 2020. doi: 10.1088/1757-899X/928/3/032019.

[24] S. A. Lifindra and I. Yuadi, “Implementation of Supervised Learning Method In Grapevine Leaf Classification,” Indonesian Journal of Artificial Intelligence and Data Mining, vol. 8, no. 1, p. 59, Nov. 2024, doi: 10.24014/ijaidm.v8i1.31633.
[25] N. I. Syahfitri, A. P. Juledi, and R. Muti’ah, “Comparative Analysis of Machine Learning Algorithm Performance in Predicting Stunting in Toddlers,” sinkron, vol. 8, no. 3, pp. 1452–1462, Jul. 2024, doi: 10.33395/sinkron.v8i3.13698.
[26] A. Hendy et al., “Supervised machine learning for classification and prediction of stunting among under-five Egyptian children,” BMC Pediatr, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12887-025-06138-x.
[27] “Ensembles: Gradient Boosting, Random Forests, Bagging, Voting, Stacking,” Scikit-learn.