Journal of Applied Data Sciences
Vol. 7, No. 2, May 2026, pp. 1614-1627
ISSN 2723-6471
1614

Performance Evaluation of Support Vector Machine (SVM) and XGBoost for Predicting Toddlers' Stunting Status Based on Anthropometric Data

Nurjoko[1], Admi Syarif[2,*], Favorisen R. Lumbanraja[3], Khairunisa Berawi[4]

[1]*Doctoral Program in Mathematics and Sciences, Lampung University, Jl. S. Brodjonegoro, Bandar Lampung, Indonesia*

[2,3]*Department of Computer Science, Faculty of Mathematics and Sciences, Lampung University, Jl. S. Brodjonegoro, Bandar Lampung, Indonesia*

[4]*Department of Medical Sciences, Faculty of Medical Science, Lampung University, Jl. S. Brodjonegoro, Bandar Lampung, Indonesia.*

(Received: November 20, 2025; Revised: January 12, 2026; Accepted: April 5, 2026; Available online: May 12, 2026)

Abstract

Stunting remains a primary global health concern, particularly in developing countries, due to its long-term effects on physical growth, cognitive development, and overall well-being. Despite various public health initiatives, challenges in early detection persist, highlighting the need for accurate, data-driven predictive models to support targeted interventions. This study aims to develop and compare the performance of two machine learning algorithms—SVM and Extreme Gradient Boosting (XGBoost)—for classifying stunting status among children under five, in order to determine the most effective method for early prediction. A quantitative machine learning approach was applied to a dataset comprising 17,498 records derived from *Posyandu* data in Lampung Province, Indonesia. The analytical pipeline included data preprocessing, class rebalancing using the Synthetic Minority Over-sampling Technique (SMOTE), and model evaluation through stratified 10-fold cross-validation. Performance was assessed using accuracy, precision, recall, and F1-score. The XGBoost model demonstrated superior performance with accuracy, precision, recall, and F1-score reaching 0.9979. In comparison, the SVM model produced slightly lower yet still strong results, achieving an accuracy of 0.9949, with similarly consistent performance across other evaluation metrics. These findings indicate that XGBoost more effectively handles high-dimensional, imbalanced data and captures nonlinear patterns in the dataset. XGBoost was identified as the optimal method for stunting classification in this study, outperforming SVM across all evaluation metrics. These results support the integration of boosting-based models into early detection systems for child nutritional assessment. Future studies should incorporate additional environmental and socioeconomic variables and evaluate model applicability in a real-time community health setting.

Keywords: Stunting, Anthropometry, SVM, XGBoost, Machine Learning, Artificial Intelligence (AI)

1. Introduction

Indonesia is one of the countries with rapid population growth, exceeding 281 million people, and a broad and diverse geographical distribution [1]. The population increase in Indonesia, which is projected to continue over the coming decades, has a significant impact on public health [2]. Based on data from the Ministry of Health, several key health indicators have shown improvement over the past decade; however, significant challenges remain in addressing complex and diverse public health issues, such as communicable diseases, non-communicable diseases, and nutritional problems, which continue to be essential concerns [3].

Stunting remains a serious public health concern in Indonesia and other developing countries. Globally, stunting is a key indicator of child nutritional status and has received significant attention within the Sustainable Development Goals framework. For example, the World Health Organization (WHO) has set a global target to reduce by 40% the proportion of children under 5 years of age experiencing stunting by 2025 [4]. Although several countries have shown a decline, the prevalence of stunting remains high in many low- and middle-income

*Corresponding author: Admi Syarif (admi.syarif@fmipa.unila.ac.id)
DOI: https://doi.org/10.47738/jads.v7i2.1260

countries, including Indonesia, indicating that stunting is a multidimensional issue requiring comprehensive, cross-sectoral approaches across health, education, sanitation, and socioeconomic sectors [5].

Stunting is a growth disorder characterized by a child's height below the expected standard for their age and sex, primarily caused by long-term nutritional deficiencies that disrupt normal linear growth, particularly during the critical first 1,000 days of life [6]. This condition is often not detected in its early stages and becomes evident only when significant delays in physical growth occur. According to the Decree of the Minister of Health of the Republic of Indonesia No. HK.01.07/MENKES/1928/2022, stunting is defined as failure to thrive due to chronic malnutrition, indicated by length or height measurements more than two standard deviations below the WHO growth standard, highlighting that stunting reflects not only inadequate food intake but also broader inequalities in meeting children's basic needs, including nutrition, health services, and environmental sanitation [7]. Furthermore, the 2024 Indonesian Nutritional Status Survey reported a national stunting prevalence of 19.8% (Ministry of Health of the Republic of Indonesia, 2024), indicating the effectiveness of government interventions such as targeted nutrition programs, improved sanitation access, maternal and child health services, and community education; however, this figure remains close to the WHO tolerance threshold of 20% for high prevalence at the population level, underscoring that stunting reduction efforts in Indonesia still require further acceleration.

The clinical practice of diagnosing stunting generally involves anthropometric examinations and biochemical analyses, with laboratory tests used to evaluate nutrient levels in children [8]. Although these methods provide a high degree of accuracy, its implementation is often constrained by the need for adequate laboratory equipment, trained personnel, and substantial financial and time investments. These challenges are particularly significant in regions with limited healthcare infrastructure. In response to these limitations, alternative methods based on Artificial Intelligence (AI) have gained increasing attention and adoption. Artificial Intelligence (AI), a field within computer science, aims to create intelligent systems that replicate human cognitive functions, such as pattern recognition, decision-making, language processing, and data-driven learning, using techniques including machine learning, fuzzy logic, artificial neural networks, and deep learning [9]. The fundamental principle of AI lies in applying computational technologies that enable machines to perceive, understand language, learn, reason, solve problems, and make decisions [10]. Among its core components, machine learning is particularly crucial, enabling systems to learn from data automatically without explicit task-specific programming [11].

Artificial Intelligence (AI) can be understood as a collection of digital technologies that enable systems or machines to learn and solve cognitive problems independently, without direct human intervention [12]. In practice, artificial intelligence is widely used for process automation, virtual agents and speech processing, predictive analytics for decision-making, sentiment analysis, and document review [13]. The focus is on two primary artificial intelligence technologies: Machine Learning (ML) and Natural Language Processing (NLP). These technologies represent key characteristics of most artificial intelligence applications in public health, as demonstrated by a cross-case analysis conducted by the European Commission and the Joint Research Centre [14].

One of the most promising areas within AI for healthcare applications is machine learning, particularly for predictive analyses using historical data [15]. Among the most widely used methods, SVM and XGBoost have demonstrated superior performance in handling large, complex datasets for predicting stunting status [16]. Both algorithms are capable of capturing nonlinear patterns in data, achieving precise classification and delivering accurate predictions. Specifically, SVM identifies an optimal hyperplane to separate classes, whereas XGBoost improves predictive performance through a boosting approach. The application of these machine learning models provides an efficient, cost-effective, and widely accessible method for the early detection of stunting using anthropometric data, particularly in resource-limited settings [17].

Currently, there is no clear consensus on which machine learning algorithm is most appropriate for predicting stunting status among toddlers using anthropometric data. Each algorithm has distinct capabilities in identifying patterns in health data, resulting in varying levels of prediction accuracy. In addition, differences in data characteristics, class imbalance, and variations in anthropometric measurement indicators present challenges that may affect prediction results. Therefore, a systematic analysis and performance evaluation of machine learning algorithms are necessary to determine the most effective method for supporting the early detection of stunting.

This research will conduct a systematic performance evaluation of various machine learning algorithms, including data preprocessing, model development and training, and model testing using comprehensive evaluation metrics such as accuracy, precision, recall, and F1-score. The evaluation aims to identify the most effective model for

ISSN 2723-6471

predicting stunting status among toddlers, thereby supporting early detection and more informed decision-making in stunting prevention efforts.

This study presents novelty through the use of routine Posyandu (Indonesian Integrated Health Service Post) data from the Lampung region, which is context-specific and rarely utilized in previous research. Unlike national survey datasets or aggregated provincial statistics that provide broad coverage but are collected periodically and often lack detailed community-level anthropometric records, the Lampung Posyandu dataset consists of routinely collected primary healthcare data with finer local granularity. This real-world dataset introduces unique characteristics, including variability in measurement practices, data incompleteness, and class imbalance patterns that reflect actual field conditions, thereby offering a more practical and contextually grounded evaluation setting. In addition, it implements an end-to-end machine learning pipeline that includes data preprocessing, SMOTE application, hyperparameter tuning, and cross-validation, as well as a systematic comparison of SVM and XGBoost algorithms to achieve more robust and accurate prediction of stunting status among toddlers. Structurally, the experimental process begins with the collection of anthropometric data on toddlers from Posyandu in the Lampung region. This is followed by data preprocessing and handling of class imbalance via SMOTE, after which the data are split into training and test sets using k-fold cross-validation. Subsequently, the SVM and XGBoost models are developed and trained through hyperparameter tuning. The final stage involves evaluating model performance using accuracy, precision, recall, and F1-score metrics to determine the most effective model for predicting stunting status among toddlers.

2. Literature Review

Previous research has shown that ML techniques, including SVM and XGBoost, are highly effective for identifying and classifying nutritional conditions, particularly stunting. For instance, a study conducted in Ghana in 2023 utilized the 2017 MICS dataset comprising 8,564 children under five to predict stunting, wasting, and underweight using seven ML algorithms. The findings revealed that XGBoost achieved the best performance, with accuracy and AUC ranging from 98% to 100%, surpassing other algorithms, including Logistic Regression, SVM, and Random Forest [18]. Similarly, a study in Zambia employed the 2018 ZDHS dataset of children under five to predict stunting using five ML algorithms, including Logistic Regression, Random Forest, SVM, XGBoost, and Naïve Bayes. The results indicated that Random Forest performed best, achieving accuracies of 79% on the test set and 61.6% on the training set. In contrast, Naïve Bayes performed the worst, reaching 61.6% on both the training and test sets [19]. These findings highlight the potential of ML applications to accelerate the diagnosis of stunting and support more timely, targeted preventive interventions.

The study conducted by [20] classified stunting status among children under five using anthropometric indicators, including sex, age, weight, height, and nutritional indices, in accordance with the CRISP-DM methodology. A total of 11 classification algorithms were evaluated, including Logistic Regression, KNN, Random Forest, AdaBoost, and Neural Networks. The results showed that the Support Vector Machine with a Radial Basis Function kernel (SVM-RBF) achieved the highest accuracy of 78% in both direct testing and 10-fold cross-validation. These findings suggest that stunting data exhibit nonlinear patterns, making kernel-based methods more effective than linear or tree-based approaches. Furthermore, a study conducted in Tunisia [21] used data from 7,963 respondents to predict the Double Burden of Malnutrition (DBM) using five machine learning algorithms. The results revealed that AdaBoost (89.8%) and SVM (89.6%) achieved the highest accuracy, whereas Naïve Bayes outperformed the other models in recall (98.1%) and AUC (91.4%). These findings highlight the effectiveness of machine learning algorithms for early detection of nutritional problems, including stunting and DBM.

The study by [22] evaluated RF, SVM, and XGBoost for stunting classification using SMOTE to address class imbalance. XGBoost combined with SMOTE achieved the best performance (accuracy 87.83%, precision 85.75%, recall 91.59%, F1-score 88.57%), outperforming RF and SVM, indicating its robustness for machine learning–based stunting detection and its potential to support public health efforts. Nevertheless, other studies have reported different outcomes [23], which compared selected machine learning algorithms for predicting child nutritional status through systematic stages of data collection, exploration, preprocessing, feature extraction, classification, and model evaluation. In this study, Random Forest achieved the best performance, with an accuracy of 0.999132, a recall of 0.999132, and an F1-score of 0.998906, making it the most consistent model for stunting detection. K-Nearest Neighbors (KNN) and Decision Trees also showed strong performance, albeit slightly lower than Random

Forest, whereas XGBoost achieved a relatively lower accuracy of 0.991033.

Research conducted by [24] examined the predictive capabilities of three machine learning methods, Random Forest (RF), SVM, and XGBoost in identifying stunting, while employing the Synthetic Minority Over-sampling Technique (SMOTE) to mitigate class imbalance. The results showed that the XGBoost–SMOTE combination delivered the highest performance, achieving an accuracy of 87.83%, precision of 85.75%, recall of 91.59%, and an F1-score of 88.57%, surpassing RF (84.56%) and SVM (68.59%). These findings suggest that coupling XGBoost with SMOTE provides an effective and reliable framework for machine learning-based stunting detection, with strong potential to enhance public health initiatives by expediting the identification of children at risk.

Previous studies on stunting prediction have employed various machine learning algorithms such as Random Forest, K-Nearest Neighbors, Decision Tree, Support Vector Machine, and gradient boosting techniques; however, their reported performance varies due to differences in dataset characteristics, feature selection, preprocessing strategies, and evaluation protocols. Some studies report the superiority of ensemble-based methods, while others emphasize the robustness of Support Vector Machine, resulting in no consistent empirical conclusion regarding the most suitable algorithm for anthropometric data. Furthermore, many studies focus primarily on reporting accuracy without consistent comparative evaluation, particularly in the context of primary healthcare service data such as Posyandu, and often overlook class imbalance handling and validation design. Therefore, this study adopts a focused comparative analysis between Support Vector Machine and Extreme Gradient Boosting, selected based on their theoretical strengths in handling high-dimensional structured data and nonlinear patterns. The evaluation design employs stratified cross-validation and a controlled oversampling procedure to produce a more reliable and transparent performance assessment.

In general, previous studies have shown that machine learning algorithms are effective for predicting stunting. However, many are limited to secondary survey data, lack an integrated analytical pipeline, and rarely account for local contexts. This study compares the performance of SVM and XGBoost using anthropometric data of toddlers from the Lampung Provincial Health Office to evaluate their effectiveness in classifying stunting status at a regional level. The findings are expected to inform the development of data-driven decision-support systems to accelerate efforts to reduce stunting in Indonesia.

3. Methodology

This section provides a comprehensive explanation of the research stages, starting with data collection and labelling, followed by data preprocessing and dataset splitting, parameter tuning to achieve optimal performance, and, finally, the implementation of classification methods and the evaluation of results as shown in figure 1.

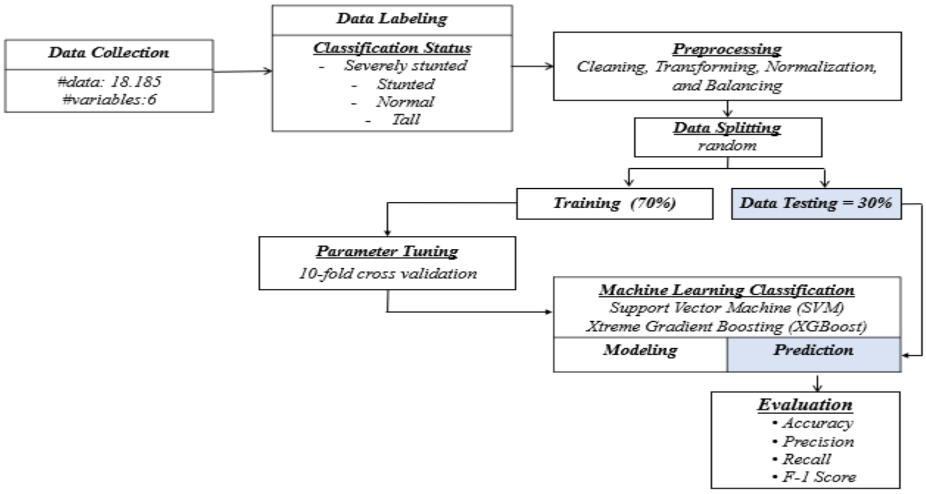


Figure 1. Research Framework and Stages

3.1. Data Collection

The dataset used in this study was obtained from the 2023 nutritional status assessment of toddlers conducted by the Lampung Provincial Health Office. It comprises 18,185 records, with eight variables: gender, date of

 ISSN 2723-6471

birth, birth weight, birth height, regency/city, current weight, current height, and date of measurement. The regency/city variable was excluded from the classification process because it does not directly contribute to the anthropometric assessment of nutritional status. Therefore, seven variables were used for model training and testing.

The target variable in this study consists of four categories—Normal, Stunted, Severely Stunted, and Tall—defined according to the WHO height-for-age z-score classification. The "Tall" category is included because it is part of the official anthropometric growth classification; however, it is not the main focus of this study, which is oriented toward stunting identification and prediction. In public health practice, primary attention is directed toward the Stunted and Severely Stunted categories, whereas the "Tall" category has limited relevance to stunting issues and contains a very small number of samples. The minimal sample size in this category may also influence model stability in the context of class imbalance.

3.2. Data Preprocessing and Splitting

Before subsequent data processing, an initial labeling stage was conducted to assign each child to a stunting category based on the height-for-age indicator. This labelling process applied the Height-for-Age Index standards for both male and female children aged 0–60 months [25]. The labelling process was carried out using the Python programming language through the following steps:

1) Data Input: Collecting child data, including gender, date of birth, height, and age.
2) Age Calculation: Determining the child's age in months.
3) Standard Table Selection: Choosing the appropriate anthropometric reference Table according to the child's gender (M for male and F for female).
4) Stunting Status Classification: Categorizing children into Severely Stunted, Stunted, Normal, or Tall groups according to the height-for-age index, with adjustments based on the child's sex.

The data preprocessing procedure was conducted using predefined and systematic criteria to ensure transparency and reproducibility. Although Microsoft Excel was used in the initial stage for data inspection and verification, all cleaning decisions were based on explicit and consistent rules. The preprocessing steps included: (1) removal of duplicate records based on unique identifiers; (2) exclusion of records with missing values in mandatory anthropometric variables such as age, weight, and height; and (3) elimination of invalid or biologically implausible entries, including zero values in key attributes and extreme outliers beyond acceptable anthropometric ranges based on WHO growth standards. Next, data transformation was performed to convert text-based attributes into numeric formats suitable for computation in Python on Google Colab. Subsequently, data normalization was applied to scale numerical features into a consistent range to optimize classification accuracy. Finally, data balancing techniques were implemented to address class imbalance, preventing the model from being biased toward majority classes and ensuring fair predictive performance across all categories.

After preprocessing, the dataset was split into two subsets: 70% for model training and 30% for testing. The training set was used to learn patterns and relationships among variables, whereas the test set was used to assess model performance on previously unseen data. This strategy supports a more impartial evaluation by measuring the model's capacity to generalize beyond the training data [26]. To obtain a more reliable and unbiased evaluation, k-fold cross-validation was implemented. This technique partitions the training dataset into k subsets of roughly equal size. In each iteration, one subset is used for validation, while the remaining subsets are employed for model training. The procedure continues until every subset has served as the validation set once. For example, with k = 10, the data are partitioned into 10 folds, with nine folds used for training and one fold for validation in each iteration. This cycle is repeated 10 times with different fold combinations, and the mean of the validation results is computed to represent the model's overall performance [27]. By combining data splitting (70:30) and k-fold cross-validation, this study ensures a more comprehensive model evaluation while minimizing the risk of overfitting and underfitting in classifying stunting status among children.

To ensure robust model evaluation and prevent information leakage, this study employed a stratified k-fold cross-validation strategy. In each fold, the dataset was divided into training and validation subsets while preserving class distribution. The Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively to the training subset after data splitting, while the validation subset remained unchanged with its original class distribution. This procedure ensured that synthetic samples were generated solely from training data and were not introduced into

ISSN 2723-6471

the validation set, thereby minimizing the risk of data leakage and avoiding overly optimistic performance estimates. Final evaluation metrics were computed by aggregating results across all validation folds to provide a more reliable assessment of model generalization ability.

Parameter adjustment aims to systematically organize model training by partitioning the dataset into multiple subsets of varying sizes to achieve optimal prediction accuracy. In this study, the data were stratified and partitioned using stratified k-fold cross-validation (CV) with k = 10. This method resamples all possible subsets of the dataset to minimize estimation bias, making it widely used in various classification studies [28].

3.3. Classification / Prediction Methods

Based on a review of related studies, this research evaluates two machine learning techniques, namely the nonlinear approaches SVM and XGBoost. ML is a branch of artificial intelligence that identifies hidden patterns in data and automatically generates predictions [29]. These advanced algorithms demonstrate superior performance compared with conventional statistical methods by effectively processing large-scale, complex, and nonlinear data [30]. In this study, the SVM and XGBoost algorithms were employed; SVM is effective at separating classes with an optimal margin for nonlinear data. At the same time, XGBoost excels in accuracy, efficiency, and handling imbalanced datasets.

The selection of SVM and XGBoost was grounded in theoretical and empirical considerations demonstrating strong performance in handling nonlinear classification problems and structured tabular data, which are consistent with the anthropometric dataset used in this study. SVM is effective in constructing optimal decision boundaries in high-dimensional spaces, while XGBoost offers robustness, regularization capability, and efficiency in modeling complex feature interactions. Although other algorithms, such as Random Forest and K-Nearest Neighbors, have shown competitive results in previous studies, this research deliberately focuses on a comparative analysis of two conceptually distinct and high-performing approaches to provide methodological depth rather than expanding the number of algorithms evaluated. Future research may broaden the scope by incorporating additional models for a more comprehensive comparison

SVM is a kernel-based supervised machine learning algorithm widely used for both regression and classification tasks. This technique constructs an optimal hyperplane that separates the training data by class label, maximizing the margin between classes [31]. The main strength of SVM lies in its ability to handle high-dimensional data and nonlinear feature spaces. Moreover, SVMs are known for producing robust models that are resistant to overfitting and deliver strong predictive performance.

During SVM training, two optimization problems, namely the primal and dual forms, are solved [32]. The two optimization formulations are shown below in primal form.

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i \tag{1}$$

As with other optimization problems, the primal problem is solved subject to the condition specified in Equation (2).

$$y_i(w \cdot x_i + b) \geq 1 - \xi i, \geq 0, i = 1, .., n \tag{2}$$

The dual form;

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j}^{n} = \alpha_i \alpha_j y_i y_j K(x_i x_j) \tag{3}$$

The solution to the dual optimization problem is obtained under the condition stated in Equation (4) below.

$$\sum_{i=1}^{n} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, n \tag{4}$$

After solving the two optimization problems mentioned above, the final model is obtained as presented in Equation (5) below.

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i x_j) + b \tag{5}$$

For a new input feature x (test set), the model predicts the class label (Positive or Negative) based on the sign of the function f(x), as presented in Equation (6) [33].

$$Predicted\ Class = sign(\sum_{i-1}^{n} \alpha_i\, y_i K(x_i x_i) + b) \tag{6}$$

For the SVM, let x denote the input feature vector (anthropometric variables), and y represent the corresponding class label (Normal, Stunted, Severely Stunted, and Tall). The objective of SVM is to determine the optimal hyperplane defined by the weight vector. $w$and bias b to maximize the margin between classes while minimizing classification error, controlled by the regularization parameter C. Since the problem is multiclass, a one-versus-rest approach is employed, where four binary classifiers are trained, and the final class is determined based on the highest decision function value.

XGBoost is widely recognized for its effectiveness in preventing overfitting, owing to a simplified objective function that combines a loss term with a regularization term. The formulation of this regularized optimization objective is expressed in the following Equation (7):

$$Obj = \sum_{m}^{n} l\,(y_{m,}\hat{y}_m) + \sum_{k}^{K} \Omega\,(f_k) \tag{7}$$

In this context, the loss function represents the discrepancy between the actual and predicted outputs. The term refers to the regularization component, defined by the following Equation (8):

$$\Omega\,(f)\; = {}_{\gamma}T + \tfrac{1}{2}\lambda\, \sum_{i=1}^{T} w^2 \tag{8}$$

In this formulation, T Denotes the number of terminal nodes (leaves) in the tree, while wRepresents the value assigned to each leaf. The terms γ and λ Regularisation parameters help manage model complexity. In this study, XGBoost was implemented as a multiclass classification algorithm to predict toddlers' stunting status. The objective function was configured using a multiclass classification scheme, where the model learns to classify each observation into one of four predefined categories (Normal, Stunted, Severely Stunted, and Tall).

XGBoost is constructed as an ensemble of decision trees, where Tdenotes the number of terminal nodes (leaves) in each tree and $w$represents the prediction score assigned to each leaf. The regularization parameters $\gamma$and $\lambda$are incorporated to control tree complexity and prevent overfitting by penalizing excessive model growth. In this study, XGBoost is implemented using a multiclass softmax objective function to address the four target categories (Normal, Stunted, Severely Stunted, and Tall). This configuration enables the model to estimate class probabilities for each observation, with the final predicted class determined based on the highest probability value.

3.4. Evaluation

In this study, model performance was evaluated using standard metrics: accuracy, precision, recall, and F1-score. These measures were calculated, analyzed, and presented using Python-based statistical and visualization libraries, thereby highlighting Python's versatility and effectiveness throughout the machine learning process.

Accuracy represents a basic indicator of a predictive model's performance, defined as the proportion of correctly predicted cases relative to the total number of observations. In this study, high accuracy was achieved across multiple ML methods by applying systematic feature selection and k-fold cross-validation. These steps were implemented to enhance the reliability and precision of the model evaluation.

$$accuracy\ = \frac{TP+TN}{(TP+TN+FP+FN)} \tag{9}$$

Sensitivity, or recall, measures a model's ability to correctly identify true positives in a dataset. It is computed as the ratio of true positives to the total number of actual positive cases. This metric reflects the model's ability to minimize missed positive detections, since undetected positive instances are classified as false negatives. Hence, sensitivity can also be expressed as the false-negative rate, which indicates the extent to which a model fails to detect existing positive cases.

$$Sensitivity\ = \frac{TP}{(TP+FN)} \tag{10}$$

Conversely, instances in which negative cases are misclassified as positive are known as false positives. Thus, specificity can also be expressed as the false-positive rate, which quantifies the proportion of negative cases misclassified as positive. A higher specificity indicates a lower likelihood that the model will produce false positives.

$$Sepecifity = \frac{TN}{(TN+FP)} \tag{11}$$

The F1-score is a performance measure that combines precision and recall via their harmonic mean, providing a more balanced assessment of a model's effectiveness. This metric is handy for imbalanced datasets because it reflects the trade-off between precision and sensitivity. Its values range from 0 to 1, where higher scores indicate stronger classification performance. An F1-score of 1 reflects flawless prediction capability, whereas a score approaching 0 indicates poor predictive accuracy. This metric is essential in scenarios where reducing both false positives and false negatives is equally important.

$$F1 - Score = 2 \text{ x } \frac{Precision \text{ x } Recall}{Precision + Recall} \tag{12}$$

In this study, precision, recall, and F1-score were computed using the weighted average method to accommodate class imbalance in the multiclass setting. This approach calculates the metric for each class and then aggregates them based on the proportion of each class in the dataset, resulting in a more representative and realistic performance evaluation, particularly for categories with unequal sample sizes.

4. Results and Discussion

The data used in this research were obtained from child growth-monitoring programs conducted at community health centers overseen by the Lampung Provincial Health Office in 2023. It consists of 18,185 records with a relatively balanced gender distribution: 10,356 boys and 7,828 girls. This distribution helps minimize potential bias in the classification process. The dataset comprises seven primary variables: sex, date of birth, birth weight, birth length, body weight at the time of assessment, and body height at the time of the evaluation. An overview of these dataset characteristics is provided in table 1.

Table 1. Research Dataset.

No	Gender	Date of birth	birth weight	birth length/height	weight measurement	height measurement
1	M	11/2/2018	3.2	45	11.6	89.9
2	F	3/7/2019	3	50	13.1	91.3
3	M	1/7/2019	3	48	13	97
…..	...	...	...	...	...	...
17496	M	4/8/2023	3.2	50	7.9	72
17497	M	7/18/2023	3	49	5.4	58
17498	M	10/24/2023	3.8	50	3.8	50

The first step of preprocessing was data cleaning, which involved identifying and removing incomplete, duplicate, and invalid entries. Numerical variables such as height and weight were validated to ensure completeness and accuracy. Records were considered invalid if key attributes, particularly weight or height, contained zero values, which typically indicate recording or input errors. The results of the data cleaning process are presented in table 2.

Table 2. Results of Data Cleaning.

Invalid Data Category	Number of Records	Description
Data with zero (0) values	548	Found in key attributes such as weight, height, or other fields that logically cannot have a zero value.
Missing data (missing values)	139	Entries lacking values for mandatory attributes are unsuitable for further processing.
Total data removed	687	Total number of records eliminated from the dataset.
Initial dataset total	18,185	Before the cleaning process.
Final dataset total	17,498	After the cleaning process.

The next stage was dataset labelling to accurately classify each child into one of four stunting categories: Normal, Stunted, Severely Stunted, or Tall. This labelling process ensured a well-structured dataset that is ready for robust and valid classification modelling. The distribution of each category in the labelled dataset is presented in table 3.

Table 3. Stunting Categories

Status Stunted	Total
Normal	16457
Stunted	866
Severely stunted	167
Tall	8

Data transformation and normalization were performed to prepare the dataset for classification by generating new attributes such as age in months, removing irrelevant variables, simplifying attribute names, and converting categorical variables into numeric formats (e.g., gender encoded as 1 for male and 0 for female, and stunting status encoded as 1 (Normal), 2 (Stunted), 3 (Severely Stunted), and 4 (Tall). The transformed dataset was analyzed in Google Colab using Python, followed by normalization with a Min–Max Scaler to standardize all numerical attributes to the 0–1 range, this step reduced bias from differences in feature scales and ensuring a proportional contribution of each attribute. These processes resulted in a cleaner, more structured dataset optimally prepared for developing SVM and XGBoost classification models, with the normalization results presented in table 4.

Table 4. Dataset Normalization

No	Gender	Birth_Weight	Birth_Height	Measured_Weight	Measured_Height	Age_(Months)
1	1	0.566667	0.727273	0.378906	0.595491	0.983333
2	0	0.500000	0.818182	0.437500	0.614058	0.916667
3	1	0.500000	0.781818	0.433594	0.689655	0.950000
….	…	...	...	...	...	...
17496	1	0.566667	0.818182	0.234375	0.358090	0.100000
17497	1	0.5000000	0.800000	0.136719	0.172414	0.050000
17498	1	0.766667	0.818182	0.074219	0.066313	0.000000

After preprocessing, the datates was split into 70% training (12,248 records) and 30% testing (5,250 records), with the training data exhibiting significant class imbalance, dominated by the Normal class. To address this issue, SMOTE was applied to generate synthetic samples for minority classes, increasing the training dataset to 46,076 records with a balanced class distribution. Furthermore, 10-fold cross-validation was implemented, with each fold using 90% of the data for training and 10% for validation, ensuring proportional representation of all classes, as shown in table 5. This approach reduced bias from a single data split, improved model generalization, and enhanced the reliability of stunting classification results.

Table 5. Training & Testing Data (Oversampled)

Status Stunted	Before	After SMOTE	Training	Testing
Normal	11519	11519	10367	1152
Stunted	606	11519	10367	1152
Severely stunted	117	11519	10367	1152
Tall	6	11519	10367	1152
Total	12248	46076	41468	4608

The modelling process for classifying stunting status among toddlers was conducted through a series of systematic experiments to evaluate model performance under different parameter configurations. The evaluation began by testing various kernel types—linear, polynomial, RBF, and sigmoid—to determine the model's ability to handle nonlinear data patterns. Furthermore, hyperparameter tuning was performed on several key parameters, including cost (C), gamma, and the maximum number of iterations, to improve classification accuracy. Model performance was assessed using accuracy, precision, recall, and F1-score as metrics. A comprehensive summary of the SVM performance across different parameter configurations is presented in table 6.

Table 6. Performance Evaluation of SVM

Parameter	Value	Accuracy	Precision	Recall	F1-Score
Kernel	Polynomial	0.9693	0.9703	0.9693	0.9693
Cost	1000	0.9935	0.9935	0.9934	0.9934
Gamma	1	0.9343	0.9407	0.9343	0.9374
Max Iteration	10000	0.9457	0.9500	0.9456	0.9477

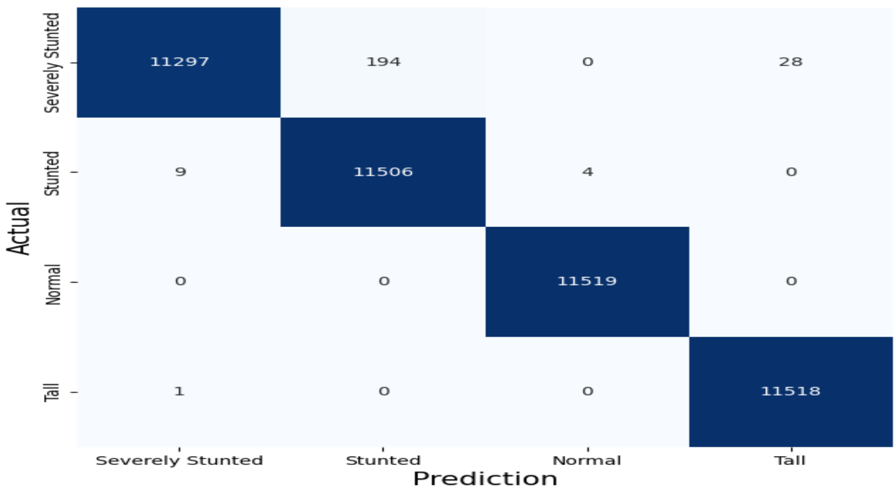
Modelling using the XGBoost algorithm was conducted across various test scenarios by adjusting key parameters, including learning rate, maximum depth, gamma, min_child_weight, subsample, and colsample_bytree. The objective was to evaluate the effect of each hyperparameter on classification performance and determine the optimal configuration. The model's performance was assessed using a confusion matrix (CM) and metrics including accuracy, precision, recall, and F1-score. A comparison of the XGBoost model performance across the tested hyperparameter configurations is presented in table 7.

Table 7. Performance Evaluation of XGBoost

Hyper Parameter	Value	Accuracy	Precision	Recall	F1-Score
Learning rate	0.9	0.9979	0.9979	0.9979	0.9979
Max depth	9	0.9978	0.9978	0.9978	0.9978
Gamma	0.1	0.9968	0.9967	0.9967	0.9967
Min child weight	1	0.9967	0.9968	0.9967	0.9967
Subsample	1	0.9966	0.9968	0.9967	0.9967
Col sample by tree	1	0.9965	0.9968	0.9967	0.9967

The performance model was evaluated using a confusion matrix (CM), which summarizes the number of correct and incorrect classifications across each class. Furthermore, performance metrics, including accuracy, precision, recall, and F1-score, were calculated to provide a comprehensive assessment of the model's predictive performance.

Based on the confusion matrix in figure 2, the classification model demonstrates excellent performance in distinguishing children's nutritional status across the four categories (Severely Stunted, Stunted, Normal, and Tall). This is reflected in the dominance of values along the main diagonal of the matrix, indicating correct predictions for each class—for example, 11,297 Severely Stunted cases, 11,506 Stunted cases, 11,519 Normal cases, and 11,518 Tall cases correctly classified. Misclassifications are relatively minimal compared to the total dataset. Some errors occur between clinically adjacent categories, such as 194 Severely Stunted cases predicted as Stunted and 28 cases predicted as Tall, as well as a small number of Stunted cases predicted as Severely Stunted (9 cases) or Normal (4 cases). This pattern indicates slight anthropometric overlap between categories with narrow z-score threshold differences.



ISSN 2723-6471

Figure 2. Confusion Matrix Model for Stunting Classification

A comparative analysis was conducted between the SVM and XGBoost models using their respective best-performing hyperparameter configurations. The SVM achieved optimal performance with a cost parameter of C = 1000, while XGBoost achieved its highest performance with a learning rate of 0.9. The best results from both models were then compared to evaluate the overall effectiveness of the classification methods, as summarized in table 8.

Table 8. Comparison of SVM and XGBoost Performance

Methode	Accuracy	Recall	Precision	F1-Score
SVM	0.9949	0.9949	0.9948	0.9949
XGBoost	0.9979	0.9979	0.9979	0.9979

As shown in table 8, XGBoost outperforms SVM across all evaluation metrics—accuracy, precision, recall, and F1-score—achieving the highest value of 0.9979, while SVM also demonstrates strong performance with slightly lower accuracy and F1-score of 0.9949. These findings suggest that XGBoost is more effective in handling complex data patterns and provides more accurate and robust stunting classifications.

However, the very high accuracy values should be interpreted cautiously, as they may indicate potential overfitting, particularly because the anthropometric variables used are directly related to the clinical definition of stunting. Although SMOTE was applied exclusively to the training data within each cross-validation fold to minimize the risk of data leakage, oversampling techniques may still lead to optimistic performance estimates. Moreover, the use of data from a single region limits the generalizability of the findings, highlighting the need for external validation using independent datasets. In addition to high overall accuracy, it is important to consider the practical implications of misclassification, particularly between the Stunted and Severely Stunted categories, which require different levels of intervention. Errors in distinguishing these categories may lead to delayed treatment or inappropriate allocation of resources. Therefore, model evaluation should not rely solely on overall accuracy but must also consider class-specific performance and its impact on public health decision-making.

5. Conclusion

Based on the performance analysis, both SVM and XGBoost demonstrated excellent capability in classifying stunting status among children under five. SVM achieved its highest accuracy of 0.9949 after hyperparameter tuning, particularly with a cost value of 1000, although optimal performance required careful adjustment of the kernel, gamma, and maximum iteration parameters. Confusion matrix evaluation showed that SVM classified most instances correctly across all classes, with minor misclassifications occurring mainly between the Normal and Stunted categories. Nevertheless, XGBoost outperformed SVM in terms of both predictive accuracy and computational efficiency, achieving a peak accuracy of 0.9979 with a learning rate of 0.9, while precision, recall, and F1-score values consistently indicated near-perfect performance. Overall, considering accuracy, efficiency, and evaluation metrics, XGBoost appears to be a more optimal approach for stunting classification, particularly for large-scale datasets and intelligent public health systems. However, despite the very high predictive performance, several limitations must be acknowledged. The dataset exhibits significant class imbalance, which may affect performance estimation even with oversampling techniques applied. Moreover, the model relies solely on anthropometric variables without incorporating socioeconomic or environmental determinants, limiting its ability to fully capture the multidimensional nature of stunting. The use of routinely collected primary healthcare data may also introduce bias due to measurement variability and incomplete records. Therefore, the findings should be interpreted cautiously in terms of generalizability, and future research is recommended to integrate broader variables and conduct external validation using independent datasets.

6. Declarations

6.1 Author Contributions

Conceptualization: N., A.S., F.R.L., and K.B.; Methodology: N., A.S., and F.R.L.; Software: N. and A.S.; Validation: N., A.S., and K.B.; Formal Analysis: N., A.S., and F.R.L.; Investigation: N. and F.R.L.; Resources: A.S. and K.B.; Data Curation: A.S.; Writing Original Draft Preparation: N., A.S., and F.R.L.; Writing Review and

ISSN 2723-6471

Editing: A.S., N., F.R.L., and K.B.; Visualization: N. and K.B.; All authors have read and agreed to the published version of the manuscript.

6.2 Data Availability Statement

The data presented in this study are available from the corresponding author upon request.

6.3 Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4 Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

This study utilized secondary data derived from routine Posyandu records. All data were fully anonymized before analysis by removing any information that could identify individuals, and the dataset was analyzed in aggregated form to ensure that no child could be personally identified. The use of data was conducted with authorization from the relevant health authorities and adhered to ethical principles for secondary data research, including confidentiality and data protection. As the study did not involve direct interaction with human subjects or access to personal identifiers, it was categorized as minimal risk research.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Imam, Statistical Yearbook of Indonesia 2024, 1st ed., vol. 52, Jakarta: BPS-Statistics Indonesia, 2024, pp. 12. [Online]. Available: https://www.bps.go.id/en/publication/2024/02/28/c1bacde03256343b2bf769b0/statistical-yearbook-of-indonesia-2024

[2] S. L. Zainalridho and S. A. Yusuf, "Population composition dynamics: a potential impact of COVID-19 pandemic on demography in Indonesia," *Populasi*, vol. 28, no. 2, pp. 54–69, 2021, doi: 10.22146/jp.63351.

[3] K. Bagchi, Towards a National Nutrition Policy: Guidelines for Countries of the Eastern Mediterranean Region, vol. 2, no. 17, Alexandria: World Health Organization, 1990, pp. 65. [Online]. Available: https://www.ncbi.nlm.nih.gov/nlmcatalog/9416927

[4] M. de Onis, K. G. Dewey, E. Borghi, A. W. Onyango, M. Blössner, and F. Branca, "The World Health Organization's global target for reducing childhood stunting by 2025: rationale and proposed actions," *Matern. Child. Nutr*., vol. 9, no. 2, pp. 6–26, 2013.

[5] T. Siswati, S. Iskandar, N. Pramestuti, J. Raharjo, A. K. Rubaya, and B. S. Wiratama, "Can the target of 40% stunting reduction by 2024 be reached?" *Iran. J. Public Health*, vol. 51, no. 10, pp. 2375–2387, 2022, doi: 10.18502/ijph.v51i10.11002.

[6] C. G. Victora, P. Christian, L. P. Vidaletti, G. Gatica-Domínguez, P. Menon, and R. E. Black, "Revisiting maternal and child undernutrition in low- and middle-income countries: variable progress towards an unfinished agenda," *Lancet*, vol. 397, no. 10282, pp. 1388–1399, 2021, doi: 10.1016/S0140-6736(21)00394-9.

[7] Kementerian Kesehatan Republik Indonesia, "Keputusan Menteri Kesehatan Nomor HK.01.07/MENKES/1928/2022 tentang pedoman nasional pelayanan kedokteran tata laksana stunting," 2022. [Online]. Available: https://peraturan.bpk.go.id/Details/280046/keputusan-menkes-no-hk0107menkes19282022

[8] K. Casadei and J. Kiel, "Anthropometric measurement," 3rd ed., vol. 2, Treasure Island (FL): StatPearls Publishing, 2025, pp. 68–72. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK537315/

[9] B. Burak, Artificial Intelligence Definitions, 2nd ed., vol. 1, Stanford: HAI (Stanford Institute for Human-Centered Artificial Intelligence), 2024, pp. 23–27.

[10] V. Bellini, M. Cascella, M. Russo, R. Lanza, and E. Bignami, "Understanding basic principles of artificial intelligence: a practical guide for intensivists," *Acta Bio Medica*, vol. 93, no. 5, pp. 1–15, 2022, doi: 10.23750/abm.v93i5.13626.

[11] H. Iqbal and S. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Comput. Sci*., vol. 2, no. 13, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[12] B. Raharja, A. P. Angga, and T. P. Analekta, "Artificial intelligence for diagnosing child stunting: a systematic literature review," *J. Syst. Manag. Sci*., vol. 13, no. 6, pp. 605–621, 2023, doi: 10.33168/JSMS.2023.0635.

[13] A. Ojo, S. Mellouli, and F. Ahmadi Zeleti, "A realist perspective on AI-era public management," in *Proc. 20th Annu. Int. Conf. Digital Government Research, Dubai*, UAE: ACM, 2019, pp. 159–170.

[14] R. Madan and M. Ashok, "AI adoption and diffusion in public administration: a systematic literature review and future research agenda," *Gov. Inf. Q*., vol. 40, no. 1, pp. 1–18, 2023, doi: 10.1016/j.giq.2022.101774.

[15] D. Dixon, "Unveiling the influence of AI predictive analytics on patient outcomes: a comprehensive narrative review," *Cureus*, vol. 16, no. 5, pp. 1–16, 2024, doi: 10.7759/cureus.59954.

[16] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, "Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia," *Public Health Nutr*., vol. 25, no. 2, pp. 269–280, 2021, doi: 10.1017/S1368980021004262.

[17] H. Shen, H. Zhao, and Y. Jiang, "Machine learning algorithms for predicting stunting among under-five children in Papua New Guinea," *Children*, vol. 10, no. 10, pp. 1–18, 2023, doi: 10.3390/children10101638.

[18] E. Anku and H. Duah, "Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms," *PLOS ONE*, vol. 19, no. 6, pp. 1–16, 2024, doi: 10.1371/journal.pone.0296625.

[19] O. N. Chilyabanyama, R. Chilengi, M. Simuyandi, C. C. Chisenga, and K. Hamusonde, "Performance of machine learning classifiers in classifying stunting among under-five children in Zambia," *Children*, vol. 9, no. 7, pp. 1–11, 2022, doi: 10.3390/children9071082.

[20] S. Syahrial, R. Ilham, Z. Asikin, and St. S. Nurdin, "Stunting classification in children's measurement data using machine learning models," *J. Multiapp*, vol. 3, no. 2, pp. 52–60, 2022, doi: 10.37899/journallamultiapp.v3i2.614.

[21] M. Zribi, F. Zaier, N. Zoghlami, J. Elati, P. Traissac, and H. Aounallah-Skhiri, "Machine learning algorithms for prediction of double burden of malnutrition among Tunisian adults," *Eur. J. Public Health*, vol. 33, no. 2, pp. 512–524, 2023, doi: 10.1093/eurpub/ckad160.1286.

[22] M. A. Hamid and E. R. Subhiyakto, "Performance comparison of random forest, SVM, and XGBoost algorithms with SMOTE for stunting prediction," *J. Appl. Inform. Comput*., vol. 9, no. 4, pp. 1163–1169, 2025, doi: 10.30871/jaic.v9i4.9701.

[23] Z. I. Bimawan, T. Astuti, and P. Arsi, "Comparison of random forest, k-nearest neighbor, decision tree, and XGBoost algorithms for detecting stunting in toddlers," *J. Tek. Inform. Jutif*, vol. 5, no. 6, pp. 1599–1607, 2024, doi: 10.52436/1.jutif.2024.5.6.2629.

[24] Z. A. A. Al-Ansi and B. M. Al-Maqaleh, "Predicting malnutrition status of under-five children in Dhamar Governorate, Yemen using data mining techniques," *Asian J. Res. Comput. Sci.,* vol. 14, no. 4, pp. 107–118, 2022, doi: 10.9734/ajrcos/2022/v14i4296.

[25] Kementerian Kesehatan Republik Indonesia, "Peraturan Menteri Kesehatan Nomor 2 Tahun 2020 tentang standar antropometri anak," 2020. [Online]. Available: https://jdih.kemkes.go.id/documents/peraturan-menteri-kesehatan-nomor-2-tahun-2020

[26] K. Pawluszek-Filipiak and A. Borkowski, "On the importance of train–test split ratio of datasets in automatic landslide detection by supervised classification," *Remote Sens*., vol. 12, no. 18, pp. 1–33, 2020, doi: 10.3390/rs12183054.

[27] V. Lumumba, D. Sang, G. Njoka, D. Musyimi, and Kavita, "Comparative analysis of cross-validation techniques: LOOCV, k-fold cross-validation, and repeated k-fold cross-validation in machine learning models," *Am. J. Theor. Appl. Stat.*, vol. 13, no. 5, pp. 127–137, 2024, doi: 10.11648/j.ajtas.20241305.13.

[28] A. Dasilas and A. Rigani, "Machine learning techniques in bankruptcy prediction: a systematic literature review," *Expert Syst. Appl*., vol. 225, pp. 12–31, 2024, doi: 10.1016/j.eswa.2024.124761.

ISSN 2723-6471

[29] M. Alghamdi, M. Al-Mallah, S. Keteyian, C. Brawner, J. Ehrman, and S. Sakr, “Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: the Henry Ford Exercise Testing (FIT) project,” *PLOS ONE*, vol. 12, no. 7, pp. 1–15, 2017, doi: 10.1371/journal.pone.0179805.

[30] Y. Zhao, B. C. Healy, D. Rotstein, C. R. G. Guttmann, R. Bakshi, and T. Chitnis, “Exploration of machine learning techniques in predicting multiple sclerosis disease course,” PLOS ONE, vol. 9, no. 4, pp. 1–17, 2014, doi: 10.1371/journal.pone.0174866.

[31] D. A. Pisner and D. M. Schnyer, “Support vector machine,” in Machine Learning: Methods and Applications to Brain Disorders, 3rd ed., vol. 2, Cambridge, MA: Academic Press, 2020.

[32] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, “Hyperparameter optimization for machine learning models based on Bayesian optimization,” *J. Electron. Sci. Technol.*, vol. 17, no. 3, pp. 26–40, 2019, doi: 10.11989/JEST.1674-862X.80904120.

[33] D. Muriithi, V. Lumumba, and M. Okongo, “A machine learning-based prediction of malaria occurrence in Kenya,” *Am. J. Theor. Appl. Stat.,* vol. 13, no. 4, pp. 65–72, 2024, doi: 10.11648/j.ajtas.20241304.11.