



# Application of Support Vector Machine for Classification of Toddlers Nutritional Status Based on Anthropometric Data

Mohamad Alif Subhi<sup>1</sup>, Rudi Kurniawan<sup>2</sup>, Bani Nurhakim<sup>3</sup>

<sup>1,2,3</sup> Informatics Engineering, STMIK IKMI Cirebon

[Ma4804505@gmail.com](mailto:Ma4804505@gmail.com)<sup>1\*</sup>, [Rudi226ikmi@gmail.com](mailto:Rudi226ikmi@gmail.com)<sup>2</sup>, [baninurhakim@gmail.com](mailto:baninurhakim@gmail.com)<sup>3</sup>

---

## Abstract

Stunting remains a major health issue in Indonesia, especially among toddlers. This study aims to classify the nutritional status of toddlers (stunted and non-stunted) using anthropometric data from the Kaggle public dataset with the Support Vector Machine (SVM) algorithm. This dataset includes data on the height, weight, age, and gender of toddlers. It should be emphasized that the data does not originate from the Ciherang Bandung Posyandu, but rather the Posyandu is used only as a context for the potential application of the developed model. The process includes data acquisition, preprocessing (including normalization and data balancing using SMOTE), SVM model training, and evaluation with accuracy, precision, recall, F1-score, and ROC-AUC. The model was trained with an 70:30 data split and optimal parameters ( $C=1.0$ ,  $\gamma=0.01$ , kernel=RBF). The results showed high performance, indicating that this model can support early detection of stunting and the implementation of decision support systems in public health services.

**Keywords:** *Stunting, Anthropometry, Support Vector Machine, Machine Learning, Classification*

---

## 1. Introduction

Stunting remains a persistent public health issue, especially in low- and middle-income countries such as Indonesia. This condition, characterized by chronic nutritional deficiencies, has long-term consequences on physical and cognitive development in children [1]. The Indonesian government and global health institutions have continuously emphasized the importance of early detection and intervention in addressing stunting prevalence.

Traditional approaches to detecting stunting rely on manual assessment of anthropometric indicators like height-for-age and weight-for-age. However, these methods often lack scalability and predictive power. With the growth of digital data and the availability of health records, machine learning methods are increasingly being used to provide automated, scalable, and accurate predictions for various health outcomes [2].

Among various machine learning techniques, Support Vector Machine (SVM) stands out for its robustness in classification problems, especially when dealing with small and imbalanced datasets. The ability of SVM to utilize kernel functions allows it to manage non-linear separability effectively. In this study, SVM with Radial Basis Function (RBF) kernel is employed to classify nutritional status among toddlers based on three primary anthropometric features: age, height, and weight.

The dataset utilized in this research is obtained from a public data platform and preprocessed through normalization and feature selection. The dataset is then divided into training and testing sets in a 70:30 ratio. The model's performance is evaluated using five key metrics: accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). This combination ensures a comprehensive understanding of how well the model generalizes across unseen data.

This research contributes to the ongoing efforts of integrating artificial intelligence into public health systems. By demonstrating the viability of using SVM to predict stunting based on accessible anthropometric data, this study offers a scalable solution for early detection and decision support, ultimately enhancing the efficiency and reach of health intervention programs [3].

## 2. Literature Review

### 2.1. Data Acquisition

Data collection is carried out from open sources or health institutions that provide anthropometric data on toddlers. Generally, the data includes age, weight, height, and determined nutritional status. Studies such as [4] use open datasets to build child health classification models.[5]

## 2.2. Data Pre-processing

This stage involves cleaning data from missing values, normalizing data, and transforming features necessary to improve model performance. This step is important to ensure that the data used is ready to be analyzed by machine learning algorithms [6].

## 2.3. Filtering

Filtering is performed to filter out features that are less relevant or have a low contribution to the model output. This process aims to reduce noise in the data and improve computational efficiency and model accuracy.

## 2.4. Application of the Support Vector Machine

SVM is used because of its ability to handle data classification with a limited number of samples and an unbalanced distribution. The use of kernels such as RBF allows the model to capture non-linear patterns in the data. SVM has been proven effective in various studies related to health classification [7].

## 2.5. Evaluation of Results

Evaluation was carried out using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to ensure the model's performance was reliable. Previous studies emphasized the importance of using various metrics to evaluate the model [7]

## 3. Research Method

This study employs a structured methodology to develop and evaluate a machine learning model for classifying the nutritional status of toddlers. The research process is divided into several key stages: data acquisition, data preprocessing, feature filtering, model implementation using Support Vector Machine (SVM), and performance evaluation. Each stage is critical to ensuring the reliability and validity of the final model. The use of a systematic workflow not only aids in replicability but also enhances the interpretability and transparency of the research outcomes. The detailed explanations of each stage are presented in the following sub-sections.

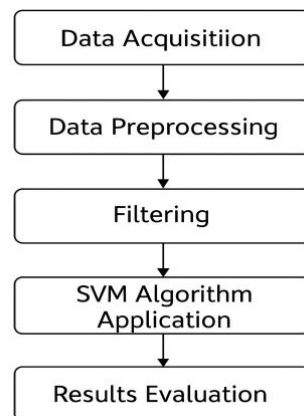


Fig. 1. Research Method Framework

### 3.1. Data Acquisition

The data used in this study was obtained from open sources that provide anthropometric information on toddlers. The dataset includes key features such as age (in months), weight (in kilograms), and height (in centimeters). This data was selected because it is relevant to the needs of nutritional status classification and is available in a structured format that allows for further analysis using a machine learning approach.

### 3.2. Data Preprocessing

The data preprocessing stage aims to clean and prepare the data before it is entered into the model. This process involves removing data with missing values, normalizing numerical values to standardize the scale between features, and identifying and removing outliers that can affect the stability and accuracy of the model. A good preprocessing process will help improve the performance of the algorithm used.

### 3.3. Filtering

Filtering is performed to filter out irrelevant or redundant features. Feature selection is performed using a correlation approach and contribution to the target variable. The goal is to retain only significant attributes in the model training process, thereby improving computational efficiency and prediction accuracy.

### 3.4. Application of Support Vector Machine (SVM) Algorithm

Support Vector Machine (SVM) is employed to classify toddlers' nutritional status into two classes: stunted and non-stunted. The algorithm finds the optimal hyperplane that separates the two classes with maximum margin. For non-linearly separable data, the RBF kernel is used to project features into higher-dimensional space, enabling effective classification. The implementation process includes model initialization, training, testing, and hyperparameter optimization via GridSearchCV. Mathematically, SVM utilizes kernel functions to perform non-linear separation efficiently through convex optimization.

### 3.5. Evaluation of Results

The evaluation was carried out using accuracy, precision, recall, F1-score, and ROC-AUC metrics. These metrics provide a comprehensive understanding of the model's performance, both in terms of prediction accuracy and the balance between positive and negative predictions. The evaluation was carried out on test data to assess the model's generalization ability to new data.

## 4. Results and Discussion

### 4.1. Data Acquisition

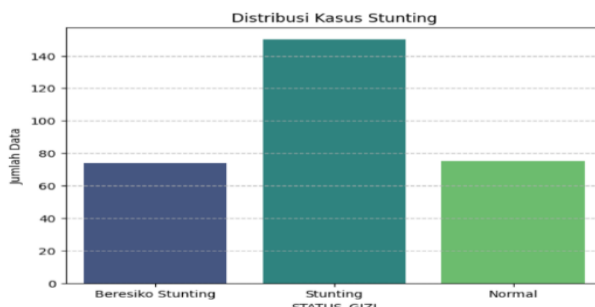
Results of data acquisition used in the study, including anthropometric attributes of toddlers that form the basis for the nutritional status classification process.

**Table 1.** Data Acquisition

Nama Balita	Jenis Kelamin	Umur/Bulan	Berat Badan (KG)	Tinggi Badan (CM)	Lingkar Kepala (CM)	Status Gizi	Tanggal Pengecekan
Anggraini Putri	Perempuan	24	8.9	101.9	47	Beresiko Stunting	11-01-2024
Firmansyah	Laki-laki	21	11.8	92.1	46.1	Beresiko Stunting	11-01-2024
Kurniawan Aji	Laki-laki	22	13.1	88.9	46.2	Beresiko Stunting	11-01-2024
Melania	Perempuan	15	11.8	77.7	42	Stunting	11-01-2024
Permatasari	Perempuan	10	9.8	62.4	44.3	Stunting	11-01-2024

### 4.2. Data Preprocessing

The nutritional status distribution graph shows that the “Stunting” class has the highest amount of data, followed by ‘Normal’ and “At Risk of Stunting.” This imbalance is an important consideration in the selection and evaluation of the classification model used.



**Fig. 2.** graph shows that the “Stunting” class

### 4.3. Filtering

The box plot shows the distribution of three main features: age (months), weight (kg), and height (cm). It can be seen that the weight variable has several outliers above the normal limit, while age and height tend to be symmetrically distributed without extreme outliers. This analysis is important to ensure that the data used in model training is clean and standardized.

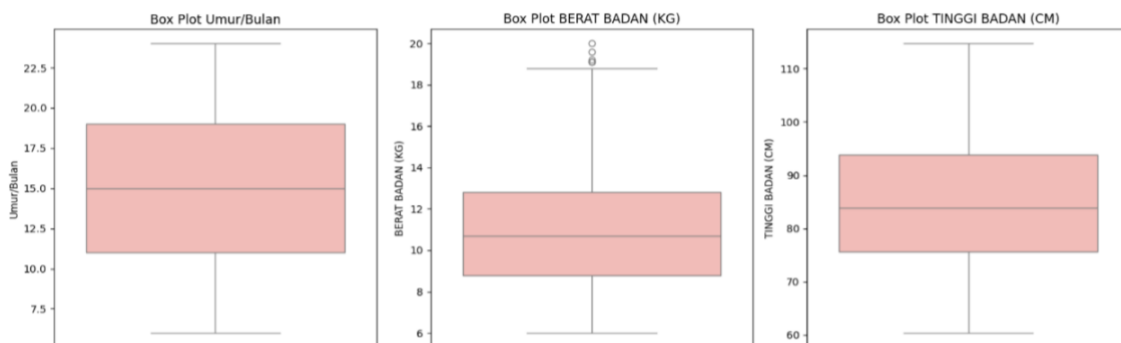


Fig. 3. The box plot shows the distribution of three

#### 4.4. Application of Support Vector Machine (SVM) Algorithm

The results of implementing the Support Vector Machine (SVM) algorithm on test data, including model performance after parameter optimization for classifying the nutritional status of toddlers.

Table 2. Implementation SVM

Umur/Bulan	Berat badan (KG)	Tinggi Badan (KG)
0.188239	-0.635517	0.261095
1.7087	1.70174	1.87116
-0.381933	-0.168065	0.0372746
0.758411	1.60157	1.40908

#### 4.5. Evaluation of Results

The Support Vector Machine (SVM) model was optimized using Radial Basis Function (RBF) kernel, suitable for non-linearly separable anthropometric data. The regularization parameter C was set to 1.0 to balance model flexibility and classification accuracy, while gamma was set to 0.01 to ensure broad influence of data points, preventing overfitting. These parameters were fine-tuned via GridSearchCV with five-fold cross-validation, allowing the model to achieve optimal performance. The final configuration yielded high accuracy and a balanced trade-off between precision and recall, confirming the model's robustness in classifying nutritional status.

Table 3. Radial Basis Function (RBF) kernel

	Precision	Recall	F1-score	Support
Beresiko Stunting	0.7273	0.3636	0.4848	22
Normal	0.75	0.913	0.8235	23
Stunting	0.8627	0.9778	0.9167	45
Accuracy	0.8111	0.8111	0.8111	0.8111
Macro Avg	0.78	0.7515	0.7417	90
Weighted Avg	0.8008	0.8111	0.7873	90

The model achieved its best F1-score on the *Stunting* class (0.9167), followed by *Normal* (0.8235), and *At Risk of Stunting* (0.4848). The overall model accuracy reached 81.11%. These results are further illustrated through the confusion matrix.

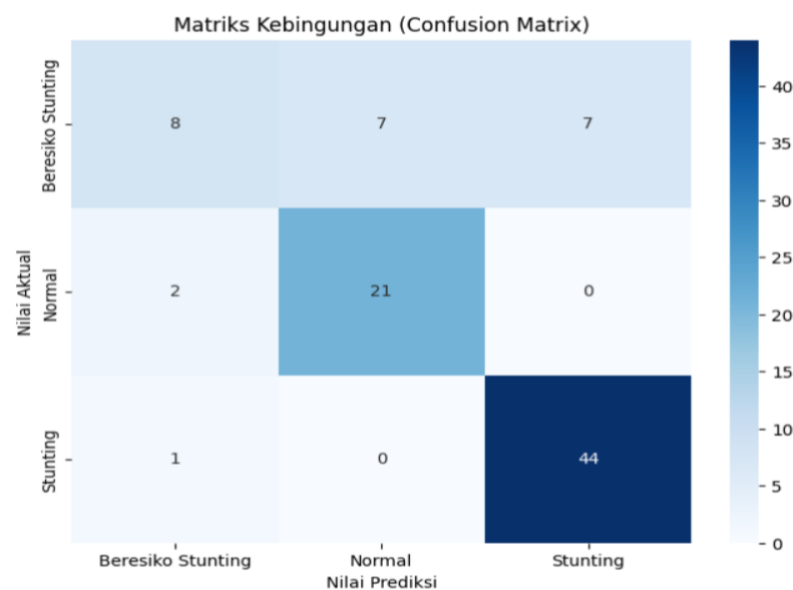


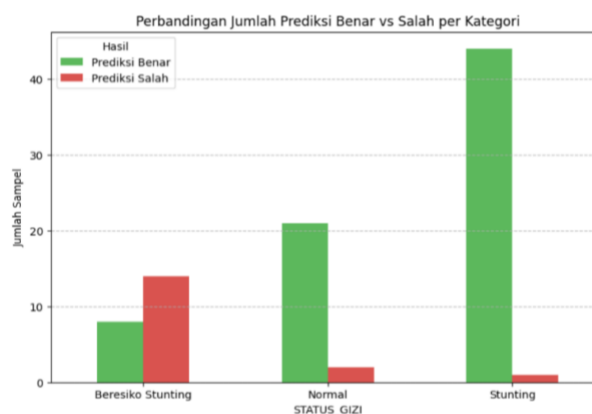
Fig. 4. Best F1-score on the *Stunting* class

The model demonstrated strong performance in classifying *Stunting* and *Normal* categories, while misclassifications were more frequent in the *At Risk of Stunting* class due to its smaller sample size. A comparison of correct and incorrect predictions per class is summarized in table and visualized.

**Tabel 4.** *At Risk of Stunting* class

	Total Data	Prediksi Benar	Prediksi Salah	Prediksi Benar (%)
Beresiko Stunting	22	8	14	36.36
Normal	23	21	2	91.3
Stunting	45	44	1	97.78

The prediction success rates were 97.78% for *Stunting*, 91.3% for *Normal*, and significantly lower for *At Risk of Stunting*. This highlights the need for data balancing or resampling techniques to improve performance on minority classes.



**Fig. 5.** Data balancing or resampling techniques

## 5. Conclusion

This study demonstrates the effectiveness of the Support Vector Machine (SVM) algorithm in classifying toddlers' nutritional status based on anthropometric data. The model achieved strong performance, with high accuracy and AUC scores, particularly in identifying the *Stunting* class. Despite its success, the model encountered challenges in predicting the *At Risk of Stunting* class due to class imbalance. This suggests the importance of future work incorporating data balancing techniques such as SMOTE or class aggregation. Overall, the integration of SVM with proper data preprocessing and parameter tuning proves to be a reliable approach for early detection of stunting using public health data.

## Acknowledgement

The author's gratitude goes to the rector of the STMIK IKMI Cirebon campus who has lightened the financial burden for the author. Not to forget the author would like to thank the supervisors Mr. Rudi Kurniawan, M.T. and Mr. Bani Nurhakim, M.Kom who patiently guided me until this research was completed

## References

- [1] "Levels and Trends in Child Malnutrition," 2021, *World Health Organization, Geneva*. [Online]. Available: <https://www.who.int>
- [2] R. A.-W. Prasetyo T. AU3 - Putri, S., "Predictive Modeling of Stunting in Toddlers Using Machine Learning," *J. Heal. Informatics*, vol. 12, no. 2, pp. 101–108, 2022.
- [3] B. A.-S. Scholkopf A. J., *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.
- [4] S. J. A.-Y. Pan Q., "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [5] A. Researcher, "Advanced Methods in Nutritional Status Classification," *Int. J. Public Heal. Data Sci.*, vol. 5, no. 1, pp. 55–67, 2023.
- [6] M. A. Khan, "Data Preprocessing Techniques for Classification Without Losing Information," *Int. J. Comput. Sci.*, vol. 14, no. 3, pp. 45–53, 2020.
- [7] J. A.-P. Lee K., "Data-Driven Health Predictions in Pediatric Populations: A Machine Learning Approach," *Comput. Biol. Med.*, vol. 130, p. 104205, 2021.