Indonesian Journal of Electronics and Instrumentation Systems (IJEIS)
Vol.x, No.x, April xxxx, pp. 1~5
ISSN (print): 2088-3714, ISSN (online): 2460-7681
DOI: 10.22146/ijeis.xxxx

Stunting Classification Model Using SMOTE and Support Vector Machine (SVM)

(Case Study: Samalanga Community Health Center)

Mahdi[1], Rahmad Hidayat[1*], Mazaia Ulfa[1]
[1]Program Studi Teknik Informatika, Jurusan Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA
Corresponding Author: rahmad_hidayat@pnl.ac.id

Abstrak

Stunting merupakan gangguan pertumbuhan yang berdampak jangka panjang terhadap perkembangan anak. Penelitian ini bertujuan mengembangkan model klasifikasi status stunting pada balita menggunakan algoritma Support Vector Machine (SVM) dengan studi kasus di Puskesmas Samalanga. Data privat diperoleh langsung dari Puskesmas Samalanga, dengan total 1.205 data balita (445 stunting dan 760 tidak stunting). Variabel yang dikumpulkan meliputi jenis kelamin, umur, berat dan tinggi badan, berat dan tinggi ideal, lingkar lengan, dan lingkar kepala. Kami juga menggunakan dataset publik yang terdiri dari variabel Sex, Age, Birth Weight, Birth Length, Body Weight, Body Length, Exclusive Breastfeeding, and Stunting. The dataset consists of 3,281 girls and 3,219 boys. untuk mengurangi ketimpangan distribusi data dan meningkatkan kinerja model pada dataset yang tidak seimbang, penelitian ini mengunakan SMOTE dengan parameter random_state=42. Sebagai tambahan, untuk mendapatkan kombinasi paramater terbaik kami menggunakan GridSearch. Model yang dikembangkan berhasil mencapai akurasi 0,97 ROC-AUC sebesar 0,96, dan f1-score rata-rata 0,97. Hasil ini menunjukkan bahwa model mampu membedakan balita stunting dan tidak stunting secara akurat. Benchmarking terhadap dataset publik menunjukkan bahwa model dalam penelitian ini memiliki akurasi lebih tinggi sebesar 2% dan nilai ROC-AUC lebih tinggi sebesar 4,7% dibandingkan penelitian sebelumnya.

Kata Kunci: Stunting, klasifikasi, Support Vector Machine, SMOTE, GridSearchCV

Abstract

Stunting is a growth disorder that has long-term impacts on child development. This study aims to develop a classification model for determining stunting status in toddlers using the Support Vector Machine (SVM) algorithm, with a case study conducted at the Samalanga Community Health Center. The private data were obtained directly from the Samalanga Community Health Center, with a total of 1,205 toddler data (445 stunted and 760 non-stunted). The variables collected include gender, age, weight, and height, ideal weight and height, arm circumference, and head circumference. We also utilized a public dataset comprising the variables Sex, Age, Birth Weight, Birth Length, Body Weight, Body Length, Exclusive Breastfeeding, and Stunting. The dataset consists of 3,281 girls and 3,219 boys. To reduce the inequality of data distribution and improve model performance on unbalanced datasets, this study used SMOTE with the random_state=42 parameter. In addition, to obtain the best parameter combination, we used GridSearch. The developed model achieved an accuracy of 0.97, an ROC-AUC of 0.96, and an average f1-score of 0.97. These results indicate that the model can accurately distinguish between stunted and non-stunted toddlers. Benchmarking against public datasets showed that

Received June 1st,2012; Revised June 25th, 2012; Accepted July 10th, 2012

the model in this study had a 2% higher accuracy and a 4.7% higher ROC-AUC value compared to previous research.

Keywords: Stunting, classification, Support Vector Machine, SMOTE, GridSearchCV

1. INTRODUCTION

Stunting is a common health problem among children in Indonesia. Stunting is a growth and development disorder caused by chronic malnutrition over a prolonged period, especially during the first 1,000 days of life [1]. Its impact is not only limited to a child's physical growth, but also increases the risk of disease, disrupts motor development, reduces productivity, and impacts the competitiveness of the nation's future generations [2]. Factors that influence stunting in toddlers include the mother's education level, the presence of infectious diseases, a history of exclusive breastfeeding, and inadequate nutritional intake, such as insufficient energy, protein, and zinc, during the growth period [3]. Poor family economic conditions also contribute to stunting, as parents may be limited in their ability to provide nutritious food for their children. Repeated infections, such as diarrhea and respiratory infections, also increase the risk of stunting because they interfere with the child's ability to absorb nutrients. In Aceh Province, the prevalence of stunted growth remains relatively high [4].

The prevalence of stunting in Aceh Province remains relatively high. According to SSGI (Indonesian Nutritional Status Survey) data, in 2022, the prevalence of stunting in Aceh Province was 31.2%. Meanwhile, the prevalence of stunting in Banda Aceh City in 2022 was 25.1%, and in Bireuen City, it was 23.4% [5]. These data indicate that, although the majority of toddlers in Aceh have a normal nutritional status, a significant stunting rate remains. Therefore, efforts to prevent and address stunting and other nutritional issues must be made to support the quality of toddler growth and development in Aceh. Currently, the process of determining stunting status in toddlers at the Samalanga Community Health Center is still limited to recording measurement results and manual interpretation by health workers based on growth charts [6]. This can cause delays in identifying toddlers at risk of stunting, especially with the large number of toddlers and limited health workers in the area. This condition presents a challenge in accelerating stunting prevention, so a machine learning-based approach is needed to assist officers [7].

As technology advances, machine learning methods can be utilized to support the automatic and accurate classification of stunting status. The Support Vector Machine (SVM) is an algorithm capable of handling both linear and non-linear classification with high accuracy [8]. The Support Vector Machine (SVM) has the advantage of detecting hidden patterns in data and is particularly suitable for classification cases, such as stunting classification. This study aims to develop a classification model for determining stunting status in toddlers using the Support Vector Machine (SVM) algorithm, with a case study conducted at the Samalanga Community Health Center. To handle data imbalance, an oversampling technique using SMOTE (Synthetic Minority Over-sampling Technique) is applied. SMOTE is employed to generate synthetic samples of minority classes, helping to balance the class distribution [9].

We build our model using two datasets. The first dataset is private data obtained directly from the Samalanga Community Health Center, comprising 1,205 toddler records (445 stunted and 760 non-stunted). The variables collected include gender, age, weight, and height, ideal weight and height, arm circumference, and head circumference. The data are then analyzed using a machine learning approach. The second dataset is a publicly available dataset obtained from [10]. This dataset contains 6,500 entries and includes variables such as Sex, Age, Birth Weight, Birth Length, Body Weight, Body Length, Exclusive Breastfeeding, and Stunting. This research is expected to contribute to the development of an accurate stunting classification method, serving as a basis for a future stunting detection system.

2. METHODS

This study aimed to develop a classification model for toddler stunting status using the Support Vector Machine (SVM) algorithm, optimized with SMOTE and GridSearchCV techniques. The architecture of the stunting classification model developed in this study is shown in Figure 1. The workflow begins when the user inputs a training dataset or new toddler data through a user-friendly application interface. All entered data, whether numerical (such as age, weight, or height) or categorical (such as gender), will be automatically saved to the database as the primary storage center. Next, the model training process is carried out using the Support Vector Machine (SVM) algorithm integrated through the Flask API. During this process, the data is trained to produce a classification model capable of identifying stunted toddlers. Once the SVM model has been trained and optimized, the system will use it to classify newly entered toddler data. The classification results are then saved back to the database and displayed on the application interface in an informative and easy-to-understand format. The Proposed Model Architecture is shown in Figure 1.

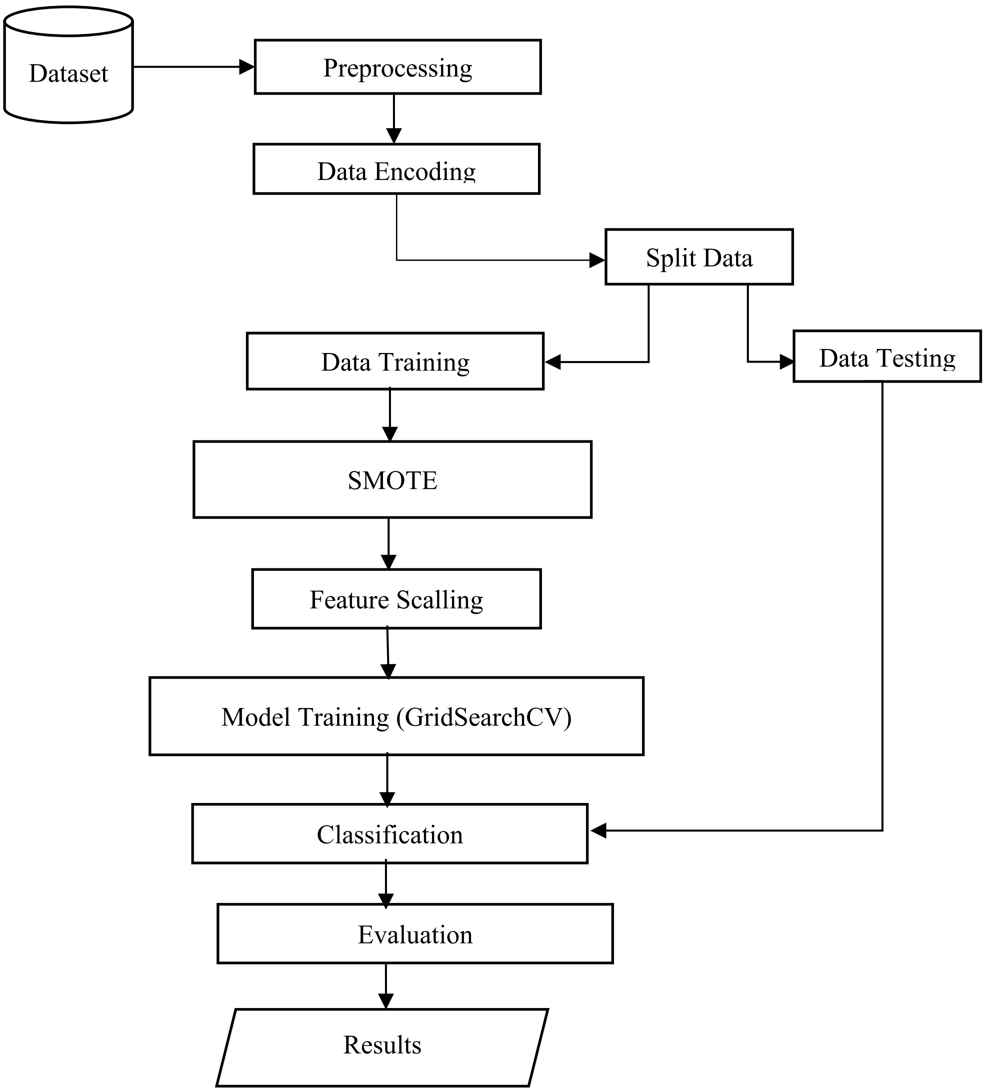


Figure 1. Proposed Model Architecture

2.1 Data Collection

This study used two types of data: private and public. The primary data was obtained from the Samalanga Community Health Center, while the public data was obtained from study by Wahyuni and Kusumodestoni (2024).

a. *Private Dataset*

Our private dataset were obtained directly from the Samalanga Community Health Center, with a total of 1,205 toddlers (445 stunted and 760 non-stunted). Variables collected included gender, age, weight, height, ideal weight, ideal height, arm circumference, and head circumference. Data collection was conducted in four stages: medical record identification, interviews with healthcare workers, initial data processing, and verification with the Community Health Center. This data was used to train and test the classification model. Table 1 shows a sample of primary data obtained from the Samalanga Community Health Center.

Tabel 1. Example data in the Private Dataset

No	Sex	Age (month)	CB (kg)	IW (kg)	CH (cm)	IH (cm)	AC (cm)	HC (cm)	Stunting
1	L	48	12.9	16.1	94.6	103.3	18	48	No
2	P	29	8.7	12.5	82.5	89.9	11	44	Yes
3	L	45	14.5	14.6	85	91.5	15	50	No
4	L	35	12.5	14.0	85	92.7	21	51	No
5	L	31	10.6	12.6	76	87.1	11	44	No
6	P	18	10.2	10.2	71	81.4	11	43.7	Yes
7	P	30	10.4	12.5	80.1	89.1	10	44.7	Yes
8	P	19	8.2	9.7	74	80.1	10	43.5	Yes
9	P	10	8.5	8.6	68	72.0	16	44.5	No
10	L	8	7.3	8.0	62.5	67.7	10	43	No

Where :
CB: Current Weight
IW: Ideal Weight
CH: Current Height
IH: Ideal Height
AC: Arm Circumference
HC: Head Circumference

b. *Public Dataset*

Public dataset comes from research conducted by [10]. The dataset consists of 6,500 entries. The variables in this dataset include Sex, Age, Birth Weight, Birth Length, Body Weight, Body Length, Exclusive Breastfeeding, and Stunting. The dataset consists of 3,281 girls and 3,219 boys. Meanwhile, the distribution based on stunting status shows that there are 3,312 children who experience stunting and 3,188 children who do not experience stunting. In this dataset, out of the total number of girls, 1,741 experience stunting, while 1,540 do not. Among the boys, 1,571 experience stunting, and 1,648 do not. This data is used for initial testing to evaluate model performance and compare it with the results of the previous study. Table 2 shows examples of data from the public dataset.

2.2 Preprocessing

Prior to model training, the data was cleaned and transformed to prepare it for use in machine learning algorithms. This process included removing missing values, detecting outliers, and adjusting data types to conform to the algorithm's required numerical format [11].

Tabel 2. Example data in the Public Dataset

No	Sex	Age (bulan)	Birth Weight (kg)	Birth Length (cm)	Body Weight (kg)	Body Length (cm)	ASI Eksklusif	Stunting
1	F	56	2.9	50	11	90	Yes	No
2	F	20	3.3	49	11.1	80.5	No	No
3	M	4	2.8	48	6.5	63	No	No
4	F	14	2.0	49	7.0	71	Yes	No
5	M	32	3.2	49	11.0	88.7	Yes	No
6	M	30	2.3	50	12.0	90.0	Yes	No
7	M	2	2.9	49	8.5	74.2	Yes	No
8	M	33	2.5	49	10.0	91.5	No	Yes
9	M	33	3.0	50	15.0	96.0	Yes	No

Categorical features, such as gender and breastfeeding status, were converted into numerical form using the label encoding method so that they could be processed by the SVM algorithm. All input features were separated from the target label (stunting status) to facilitate the training and evaluation process. The data was divided into two parts: 80% for training and 20% for testing, using the train_test_split function from the Scikit-learn library. All numerical features were standardized using StandardScaler to have a mean of zero and a standard deviation of one. This process improved the stability and convergence of the SVM model.

2.3 Data Balancing with Synthetic Minority Oversampling Technique (SMOTE)

The Synthetic Minority Oversampling Technique (SMOTE) is an advanced oversampling method that synthetically generates novel instances within the minority class, rather than merely duplicating existing samples [12]. This methodology is designed to mitigate class imbalance within datasets, thereby enhancing the performance of classification algorithms trained on such data [13], [14]. In the present study, SMOTE objects are instantiated with the parameter random_state=42 to guarantee the reproducibility of results across different executions. SMOTE is exclusively applied to the training set to achieve class balance, thereby preventing potential data leakage into the testing set. Upon completion of the SMOTE procedure, the sample size and label distribution are re-examined to confirm that class balance has been attained.

2.4 Support Vector Machine (SVM)

SVM is a classification method in machine learning (supervised learning) that predicts classes based on patterns obtained from the training process, and this method was developed by Vladimir Vapnik [15]. Support Vector Machine (SVM) is a supervised learning algorithm used to recognize patterns in data and is widely applied in classification, regression, and anomaly detection tasks . This method works by finding the best hyperplane that maximally separates data from two classes. The primary challenge of this method is selecting optimal parameters, such as C and gamma values, as well as handling imbalanced data. In this study, the optimization process was carried out using GridSearchCV, and data balancing was performed using the SMOTE method.

The stages of applying the SVM algorithm in this stunting classification model are as follows:

a. Determine the dataset to be used.
b. Perform data preprocessing, including data cleaning, normalization, and handling missing values.
c. Divide the dataset into training and test data.
d. Balance the class distribution using the Synthetic Minority Oversampling Technique (SMOTE).
e. Search for the best parameters using GridSearchCV to obtain optimal C and gamma values.
f. Train the model using the training data and the best parameters.
g. Test the model's performance on the test data using accuracy, precision, recall, F1-score, and ROC-AUC metrics.
h. Implement the model into a web-based classification system using Laravel and Python Flask for model integration.

SVM works by determining the optimal hyperplane that maximally separates two data classes. This hyperplane is determined by the support vector, which is the closest data point from each class. The Support Vector Machine (SVM) offers various types of kernel functions, including linear, RBF, polynomial, and sigmoid. The selection of this kernel type is crucial, particularly in determining the optimal feature space for effective data separation [9]. The hyperplane effectively separates positive (+1) and negative (–1) data. To calculate the class prediction for new data based on the hyperplane that has been formed, the following decision function is used [16]:

$$f((x)) = sign\ (w.\emptyset(x)+b \tag{1}$$

In certain cases, the kernel function $\emptyset(x)$ is used to map the data to a higher dimension, allowing for linear separation. Thus, the prediction function can be rewritten as:

$$f((x)) = sign\ (\textstyle\sum_{i=1}^{n} a_i y_i \emptyset(x_i)^T.\emptyset(x)+b) \tag{2}$$

In order to obtain the optimal hyperplane that separates two classes of data, SVM minimizes the following objective function:

$$Minimize \int 1[w] = \frac{1}{2}||w||^2 \tag{3}$$

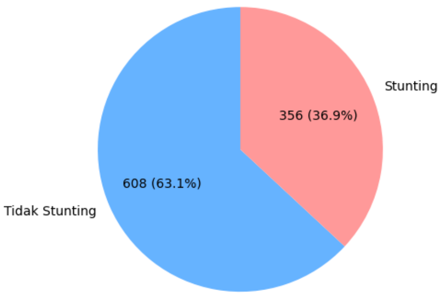
Where:
X_i: i-th data
w: vector weight
b : bias value
Y_i: i-th data class

In this study, the kernel used is the Radial Basis Function (RBF), as it is well-suited for handling non-linear data and yields high accuracy in classification. The RBF kernel uses a Gaussian function (known as a radial basis function) to measure the similarity between two input vectors in the feature space [11]. One of the frequently used kernels is the Radial Basis Function (RBF). The RBF kernel equation is:

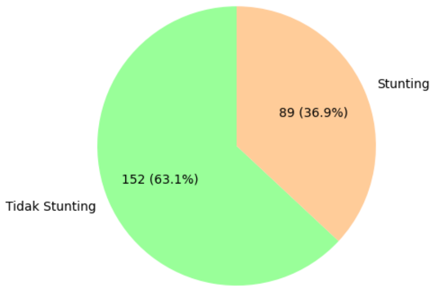
$$K(x,x') = exp\ (-\gamma(x,y)^2 \quad (4)$$

3. RESULTS AND DISCUSSION

A total of 1205 data points contained in the primary dataset were divided into training and testing data points. The training and testing data points were divided in an 80:20 ratio. Stratification was used to maintain a balanced proportion of target classes between the training and testing data points. The data points were divided into 964 for training and 241 for testing. The label distribution in the training data (Figure 2a) consisted of 608 data points labeled as "Not Stunting" (class 0) and 356 data points labeled as "Stunting" (class 1). Meanwhile, in the testing data points, there were 152 data points classified as 'Not Stunting' and 89 data points, as shown in Figure 2b.



a. Training Data



b. Testing Data

Figure 2. Splitting Data

SMOTE works by generating synthetic samples for the minority class (stunting), balancing them with the majority class (non-stunting). This step is crucial to ensure the model is unbiased and can recognize both classes fairly during training. Figure 3 below shows the distribution of training data after the SMOTE process.

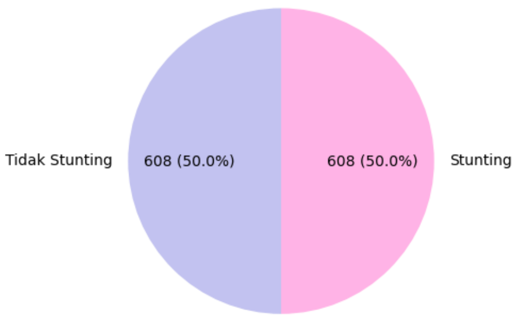


Figure 3. SMOTE result on training data

This study applied the Support Vector Machine (SVM) algorithm with a Radial Basis Function (RBF) kernel to classify stunting status in toddlers. Kernel RBF dipilih karena cocok dengan data non linear, dan berdasarkan penelitian [17] kernel RBF mengungguli kernel SVM lainnya dalam hal akurasi dan generalisasi. The model training process was carried out systematically using a pipeline that included data standardization, class balancing using SMOTE, and automatic hyperparameter tuning with GridSearch. The training dataset consisted of 964 data points, while the testing dataset consisted of 241 data points. The configuration of GridSearch is shown in Figure 4. The best model was obtained with parameters C = 10 and gamma = scale, resulting in a testing accuracy of 0.97 and an ROC AUC value of 0.96. The confusion matrix results show that of the 152 non-stunted toddlers, 150 were classified correctly (True Negative) and 2 incorrectly (False Positive).

```
# Pipeline dan Grid search
        pipeline = Pipeline([
            ("scaler", StandardScaler()),
            ("svm", SVC(kernel="rbf", probability=True, random_state=42)),
        ])
        param_grid = {
            "svm__C": [0.01, 0.1, 1, 10],
            "svm__gamma": ["scale", 0.01, 0.1],
        }
        grid = GridSearchCV(pipeline, param_grid, cv=5, n_jobs=-1)
        grid.fit(X_train_resampled, y_train_resampled)

        best_model = grid.best_estimator_
```

Figure 4. Gridsearch Configuration

Meanwhile, of the 89 stunted toddlers, 83 were classified correctly (True Positive) and 6 incorrectly (False Negative). Based on the classification report, the model has high precision, recall, and F1-score in both classes. For the "Not Stunting" class, the precision was 0.96, recall 0.99, and F1-score 0.97. Figure 5 shown the model evaluation results.

```
Training Accuracy : 98.55%
Testing Accuracy  : 96.68%
ROC AUC           : 0.9638


Classification Report:

              precision   recall   f1-score   support

Not Stunting     0.96      0.99      0.97       152
    Stunting     0.98      0.93      0.95       89
```

Figure 5. Model Evaluation Results

For the "Stunting" class, the precision was 0.98, the recall was 0.93, and the F1-score was 0.95. The average F1-score reached 0.97, indicating good model performance. The heatmap of the proposed model is shown in Figure 6. The ROC (Receiver Operating Characteristic) curve is used to evaluate the performance of the classification model in distinguishing between two classes. The horizontal axis (False Positive Rate) represents the proportion of negative data incorrectly classified as positive, while the vertical axis (True Positive Rate) shows the proportion of positive data correctly classified. The gray diagonal line shows the baseline or random performance, where the model has no ability to distinguish between classes. In the resulting curve, it can be seen that the blue line is significantly above the diagonal line, approaching the coordinate point (0, 1), which indicates excellent model performance.

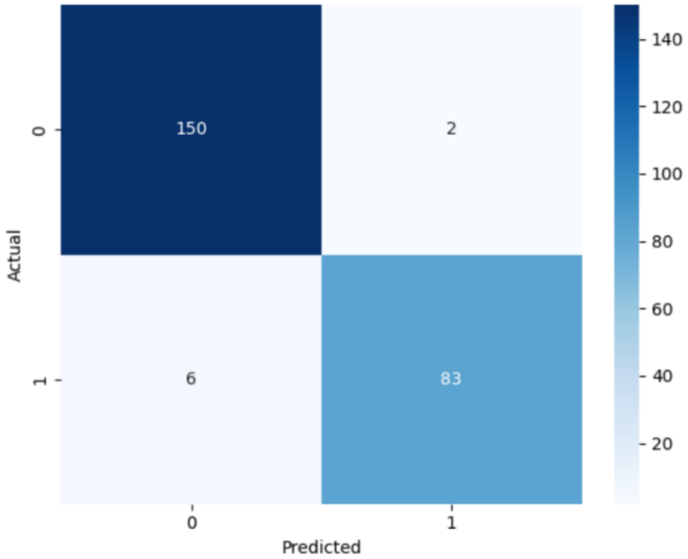


Figure 6. Proposed Model's Heatmap

The Area Under the Curve (AUC) value obtained is 0.96, which means the model has a 96% ability to distinguish between positive and negative classes. Generally, an AUC value above 0.90 is considered excellent model performance. Thus, the model built in this study shows very high performance and is reliable for the classification process. Proposhed model ROC shown in Figure 7.

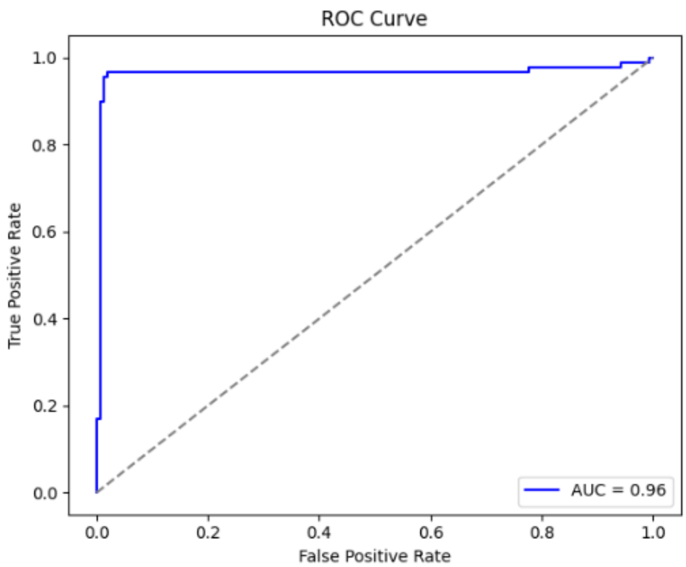


Figure 7. Proposed Model's ROC Curve

Model evaluation reveals that the proposed method offers significant advantages over the study by Wahyuni & Kusumodestoni (2024). Although both use the SVM algorithm, this study employs parameter optimization using GridSearchCV and data imbalance handling with SMOTE, which has been proven to improve model performance. Test results show an increase in accuracy from 93% to 95%, precision from 93% to 95%, recall from 93% to 95%, and F1-score from 93% to 95%. In addition, the ROC-AUC value increased from 92.6% to 97.3%, indicating a better model's ability to distinguish classes overall. The performance comparison on the public dataset is shown in Table 3.

Table 3. Benchmarking on Public Dataset

Study	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Proposed Method	0.95	0.95	0.95	0.95	0.97
Wahyuni & Kusumodestoni (2024)	0.93	0.93	0.93	0.93	0.93

4. CONCLUSIONS

This study developed a classification model for stunting status in toddlers using the Support Vector Machine (SVM) algorithm, optimized through a systematic pipeline approach that included feature standardization, data balancing with SMOTE, and hyperparameter tuning with GridSearchCV. Model evaluation results using primary data showed that the SVM model with the best parameters (C=10, gamma=scale) was able to achieve an accuracy of 0.97, an ROC-AUC of 0.97, and an average F1-score of 0.97 on test data from the Samalanga Community Health Center. These values reflect excellent classification performance with a low error rate. Furthermore, benchmarking results against previous studies using similar public datasets showed that the SVM approach in this study provided superior results. The developed model achieved an accuracy of 0.95 and an ROC-AUC of 0.97, outperforming the comparison model, which achieved an accuracy of 0.93 and an ROC-AUC of 0.93. This proves the effectiveness of integrating SMOTE and automatic tuning in improving the performance of the classification model on imbalanced data.

5. FUTURE WORKS

For future research, it is recommended that the dataset be expanded by collecting data on toddlers from various geographic regions and over a wider time span, allowing the model to have better generalizability across diverse populations. Furthermore, further research should compare the performance of the Support Vector Machine (SVM) algorithm with other classification methods, such as Random Forest, XGBoost, or Neural Networks, to identify the most optimal approach for classifying stunting. Equally important, exploration of additional medically relevant features, such as maternal health history, child nutritional intake patterns, and residential environmental conditions, is also highly recommended. The addition of these features is expected to improve the accuracy and depth of the model's analysis in identifying factors contributing to stunting risk more comprehensively and precisely.

REFERENCES

[1] G. A. F. Khansa and P. H. Gunawan, “Predicting Stunting in Toddlers Using KNN and Naïve Bayes Methods,” *2024 Int. Conf. Data Sci. Its Appl. ICoDSA 2024*, pp. 17–21, 2024, doi: 10.1109/ICoDSA62899.2024.10651676.

[2] A. T. Dhafin, R. Fauzi, and Z. F. Azzahra, "Development of Tracking Features and Recommendation for Mother's Nutrition in the Genting Mobile Application to Prevent Stunting Using Vision Transformer (Case Study: PKK Bandung City)," *2024 Int. Conf. Intell. Cybern. Technol. Appl. ICICyTA 2024*, pp. 111–116, 2024, doi: 10.1109/ICICYTA64807.2024.10913053.

[3] A. F. Kanz, L. P. Johanes, I. N. Alam, and L. A. Wulandhari, "Stunting Prediction in Children Using Random Forest Algorithm," *2024 6th Int. Conf. Cybern. Intell. Syst. ICORIS 2024*, pp. 1–4, 2024, doi: 10.1109/ICORIS63540.2024.10903911.

[4] H. Janawisuta, P. H. Gunawan, and Indwiarti, "Early Detection of Stunting in Indonesian Toddlers: A Machine Learning Approach," *2024 Int. Conf. Data Sci. Its Appl. ICoDSA 2024*, pp. 12–16, 2024, doi: 10.1109/ICoDSA62899.2024.10651637.

[5] K. K. R. Indonesia, "Laporan Hasil Survei Status Gizi Indonesia (SSGI) Tahun 2022." Accessed: Jul. 09, 2025. [Online]. Available: https://stunting.go.id/buku-saku-hasil-survei-status-gizi-indonesia-ssgi-2022/

[6] A. Yusuf, "K-Means Clustering Based on Distance Measures: Stunting Prevalence Clustering in South Kalimantan," *2022 5th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2022*, pp. 706–710, 2022, doi: 10.1109/ISRITI56927.2022.10052925.

[7] D. M. Azis, R. Fauzi, and S. Suakanto, "Development of Stunting Prediction Features To Prevent Stunting Using Support Vector Machine (Svm) Algorithm," *ICONDBTM 2024 - Proc. Int. Conf. Digit. Bus. Technol. Manag.*, pp. 1–6, 2024, doi: 10.1109/ICONDBTM64135.2024.11122841.

[8] R. Hidayat, A. Harjoko, and A. Musdholifah, "Backpack detection model using multi-scale superpixel and body-part segmentation," *Int. J. Smart Sens. Intell. Syst.*, vol. 16, no. 1, pp. 1–12, 2023, doi: 10.2478/ijssis-2023-0008.

[9] D. N. Fadhilah and P. H. Gunawan, "Support Vector Machine-Based Classification of Toddler Stunting in Bandarharjo," *COMNETSAT 2024 - IEEE Int. Conf. Commun. Networks Satell.*, pp. 263–267, 2024, doi: 10.1109/COMNETSAT63286.2024.10862872.

[10] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bull. Inf. Technol.*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.

[11] R. Hidayat, M. Khadafi, and M. Davi, "Klasifikasi Tumor Payudara Pada Citra Ultrasonografi Menggunakan Multi-fitur Tekstur dan Support Vector Machine," *J. Serambi Eng.*, vol. IX, no. 2, pp. 8805–8811, 2024.

[12] Y. Bao and S. Yang, "Two Novel SMOTE Methods for Solving Imbalanced Classification Problems," *IEEE Access*, vol. 11, no. December 2022, pp. 5816–5823, 2023, doi: 10.1109/ACCESS.2023.3236794.

[13] Y. Huang, "FS-SMOTE: An Improved SMOTE Method Based on Feature Space Scoring Mechanism for Solving Class-Imbalanced Problems," *IEEE Access*, vol. 13, no. July, pp. 148074–148082, 2025, doi: 10.1109/ACCESS.2025.3597794.

[14] R. Mohammed and E. M. Karim, "FADA-SMOTE-Ms: Fuzzy Adaptative Smote-Based Methods," *IEEE Access*, vol. 12, no. November, pp. 158742–158765, 2024, doi: 10.1109/ACCESS.2024.3480848.

[15] R. Hidayat, M. Haris, and Z. K. Simbolon, "Implementasi Metode Support Vector Machine (SVM) Pada Klasifikasi Drop Out (DO) Mahasiswa," vol. 9, pp. 51–59, 2024.

[16] T. Khanam, K. Deb, and K. H. Jo, "Baggage detection and classification using human body parameter & boosting technique," *Proc. - 2017 10th Int. Conf. Hum. Syst. Interact. HSI 2017*, no. C, pp. 54–59, 2017, doi: 10.1109/HSI.2017.8004996.

[17] S. Vanet, G. O. Anyanwu, C. I. Nwakanma, J. Lee, D. Kim, and S. Member, "Optimization of RBF-SVM Kernel Using Grid Search Algorithm for DDoS Attack Detection in," *IEEE Internet Things J.*, vol. 10, no. 10, pp. 8477–8490, 2023, doi: 10.1109/JIOT.2022.3199712.