JCSE Journal of Computing Science and Engineering

Regular Paper
Journal of Computing Science and Engineering,
Vol. 18, No. 4, December 2024, pp. 203-213

Enhancing Stunting Prediction for Indonesian Children Using Machine Learning with SMOTE Data Balancing

Aqeela Fathya Najwa and Putu Harry Gunawan*
CoE HUMIC, School of Computing, Telkom University, Bandung, Indonesia
aqeelafn@student.telkomuniversity.ac.id, phgunawan@telkomuniversity.ac.id

Abstract
Stunting is a significant health issue in Indonesia, affecting the growth of children under five due to chronic malnutrition. Traditional methods for early identification of at-risk children often fall short, highlighting the need for advanced predictive models like machine learning (ML). This study compares the performance of support vector machine (SVM) and Decision Tree algorithms using data from Bandarharjo Community Health Centers. Initial results show poor performance for SVM models with linear and polynomial kernels, achieving F1-scores between 0% and 54%. The decision tree algorithm performed slightly better with an F1-score of 64%. To improve detection accuracy, the Synthetic Minority Over-sampling Technique (SMOTE) was applied as a data balancing technique to address class imbalance. After applying SMOTE, the decision tree achieved an F1-score of 97%, proving to be the most effective model. The SVM with the radial basis function (RBF) kernel also improved significantly, achieving an F1-score of 94%. These findings demonstrate that data balancing techniques like SMOTE are crucial for enhancing the accuracy and effectiveness of ML models in detecting stunting, enabling more timely and accurate health interventions.

Category: Smart and Intelligent Computing

Keywords: Stunting; Support vector machine; Decision tree; Machine learning

I. INTRODUCTION

Stunting is a major health issue for children in Indonesia [1], characterized by the failure of children under 5 years old to grow due to chronic malnutrition during the first 1,000 days of life, resulting in stunted growth [2-4]. Addressing this issue is crucial to avoid long-term consequences for both children and the wider community and to prevent a decline in long-term health. Notably, President Joko Widodo has set a target to reduce the prevalence of stunting from 21% in 2022 to 14% in 2024, as outlined on the website (https://sehatnegeriku.kemkes.go.id/). Effective prevention of stunting involves close monitoring of children's growth, underscoring the need for a system capable of predicting stunting conditions [5-8].

To predict stunting conditions in children, appropriate algorithms are needed for classification [9]. According to research conducted in [10], the support vector machine (SVM) algorithm has a high recall value of 1.00, indicating its capability in identifying stunting. On the other hand, research in [11] found that the decision tree algorithm has a high accuracy of 96.15%, with a difference of 33.67% compared to SVM.

A comparative analysis between the SVM and decision tree algorithms needs to be conducted to validate and deepen previous findings and provide additional contributions

Open Access http://dx.doi.org/10.5626/JCSE.2024.18.4.203 http://jcse.kiise.org

Received 05 September 24; **Accepted** 12 December 24
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

through algorithm optimization for more detailed predictions [12]. This evaluation is important considering the limited feature analysis and dataset usage in previous studies.

This study aims to develop machine learning models to detect stunting conditions in children. Using a dataset from Bandarharjo Community Health Centers, this research focuses on training models to provide early warnings. Through an in-depth comparison of two machine learning models—SVM and decision tree—it is expected to identify the most effective model. This study hopes to contribute to the prevention of stunting in Indonesia by providing accurate and efficient predictions, enabling healthcare professionals to take timely necessary interventions

II. MODEL

A. Research Design

This study began by defining the research problem, which is classifying the nutritional status of children, followed by planning that included setting objectives, selecting analysis methods, and identifying data sources. A literature review was conducted to understand relevant studies and methods, including the use of algorithms such as SVM and decision tree, as well as techniques for handling imbalanced data such as SMOTE (Synthetic Minority Over-sampling Technique).

Data was collected from Bandarharjo Community Health Centers and imported from Excel files into a DataFrame using the pandas library. After importing the data, the next step was to understand the data structure, clean the data from duplicate rows, and perform distribution analysis to identify the proportions of ‘Normal’ and ‘Stunting’ categories.

Relevant features such as weight, height, and age in months were selected for analysis, and the data was then visualized to identify patterns and outliers. The data was prepared for further analysis by splitting the dataset into features and targets and dividing it into training and testing sets. The data was also standardized using StandardScaler and converted to a compatible format for analysis.

Data balance checks were performed, and if necessary, SMOTE technique was applied to balance the classes in the dataset. SVM and decision tree algorithms were applied to build classification models, with accuracy results compared using evaluation metrics such as accuracy, precision, recall, and F1-score. The best-performing model was selected for final analysis.

The research conclusions include a comparison of models and recommendations for child health interventions based on the findings. The final report was prepared to provide useful insights for Bandarharjo Community Health Centers and related parties to improve child health in the region. Fig. 1 illustrates the flowchart used in this research.

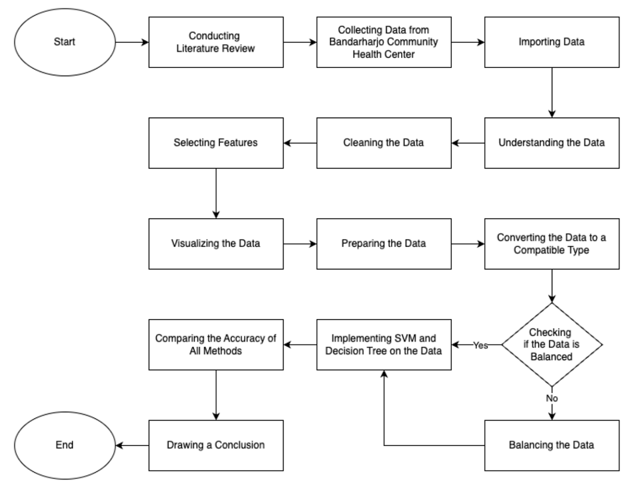

Fig. 1. Research flowchart.

B. Support Vector Machine

SVM is a machine learning technique in the supervised learning category used for classification, regression, and data prediction [13]. The main objective of SVM is to find the best hyperplane that separates two classes of data in a high-dimensional space. A hyperplane is a line or surface that separates data between classes with the maximum margin [10]. The data points closest to the hyperplane are called support vectors and determine the position of the hyperplane [11].

The SVM algorithm uses training data to develop a classification model, which is then used to predict the class of new data. SVM seeks a hyperplane that maximizes the margin between classes in the dataset [11]. The basic equation of the hyperplane in SVM is [14]:

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

where w is the weight vector, x is the feature vector, and b is the bias [14].

C. Decision Tree

Decision tree is a machine learning algorithm for predictive modeling and data analysis. This algorithm constructs a decision tree structure consisting of decision nodes connected by branches from the root node to leaf nodes, facilitating the visualization and understanding of the decision-making process [15].

The decision tree uses a set of rules to make decisions depicted as branches on the tree. The algorithm splits the dataset based on features with the highest information gain at each node. This process repeats recursively until it reaches the leaf nodes, where the final decision is made based on the majority class at that node [16].

III. RESULTS AND DISCUSSION

A. Exploratory Data Analysis

The dataset used in this study is sourced from the Bandarharjo Community Health Centers, encompassing health-related data of children under 5 years old, including variables such as weight, height, and age. After cleaning duplicate entries, the dataset consists of 3,674 records, with key features including: weight, height, age in months, stunting status, and nutritional status.

The initial step in the analysis involves examining the distribution of these features through boxplots, as shown in Fig. 2.

The analysis reveals several extreme values in the weight feature, with significant outliers at the upper end of the distribution. The distribution of TB/U z-scores also shows some outliers. For the height and age features, the data are relatively consistent without extreme outliers, suggesting that the children's height and age ranges are within reasonable bounds.

Further exploratory data analysis includes a comparison of stunting and normal status among the children, presented in Fig. 3. The analysis reveals a significant imbalance,

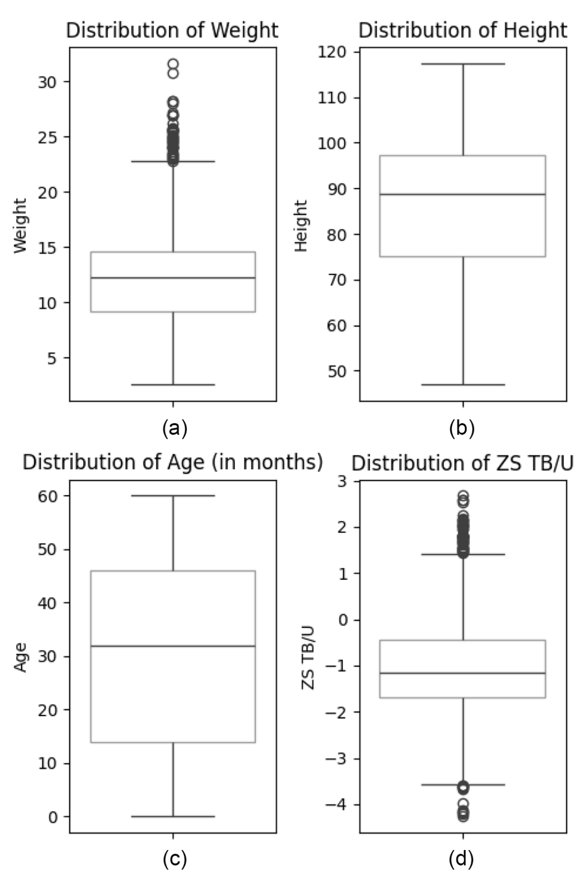

Fig. 2. Boxplot analysis of (a) weight, (b) height, (c) age, and (d) z-score TB/U.

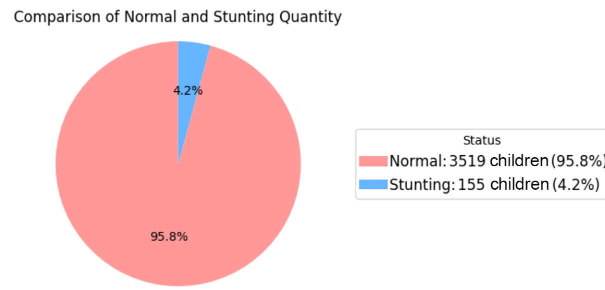

Fig. 3. Comparison of normal and stunted children.

with only 4.2% of the children classified as stunted, while the majority, 95.8%, are categorized as normal. This imbalance presents a potential challenge in analysis and modeling, as the dominant normal category can bias the results.

To delve deeper, Fig. 4 illustrates the nutritional status within the stunted group, showing a range of categories including good nutrition, malnutrition, severe malnutrition, risk of overnutrition, and overnutrition. This additional context highlights the variation in nutritional status among stunted children, which is crucial for a more comprehensive understanding of the dataset.

Given the significant imbalance between the target classes, preprocessing steps are necessary to prepare the data for modeling. The target column is then transformed from the 'TB/U' category into two binary classes: 'Normal' and 'Stunting,' to simplify the classification process.

To address the class imbalance issue, SMOTE was applied to generate synthetic samples for the minority 'Stunting' class, reducing the disparity between the 'Normal' and 'Stunting' classes [17]. As illustrated in Fig. 5, before applying SMOTE, the dataset exhibited a 1:23 ratio between the minority and majority classes. By creating synthetic data that mirrors the characteristics of the minority class based on existing features, SMOTE balances the class distribution, reducing bias towards the majority class and enabling the model to learn more effectively from both classes [18].

This preprocessing step is critical because balanced data ensures that the classification model is more accurate and fair in identifying the 'Stunting' class, thus reducing the

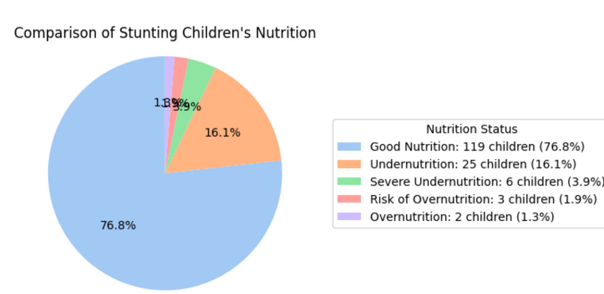

Fig. 4. Comparison of stunting children's nutritional status.

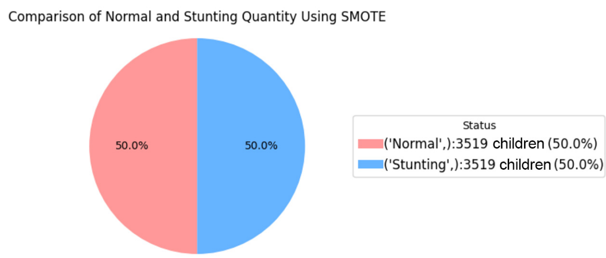

Fig. 5. Comparison of normal and stunted children using SMOTE.

likelihood of misclassification due to extreme class imbalance. After applying SMOTE, the data is split into training and testing sets, followed by normalization to maintain consistency and effectiveness in subsequent model training. This comprehensive approach to data exploration and preprocessing establishes a strong foundation for developing robust classification models.

B. Implementation SVM and Decision Tree

1) SVM Implementation without Balanced Parameter

In this study, SVM was implemented to classify stunting data using various kernel types to determine the optimal performance. The implementation process was carried out using the scikit-learn library in Python. The dataset was split into training and testing sets with an 80:20 ratio, and data scaling was performed into the range [0,1]. Four types of kernels were tested: linear, polynomial, radial basis function (RBF), and sigmoid [19]. The evaluation results for each kernel are presented in Fig. 6. This method identifies how each kernel adapts to the stunting data, including its ability to detect the minority class 'Stunting."

From the analysis of the confusion matrix across various SVM kernels, as shown in Fig. 6, there are significant differences in stunting detection ability for each kernel. The linear and polynomial kernels show unsatisfactory results in detecting stunting, with very low numbers of stunting detections (0 for linear and 1 for polynomial), despite having high accuracy in detecting normal cases. In contrast, the RBF kernel demonstrates better performance with 10 stunting cases detected, although there are still some undetected cases. The Sigmoid kernel also shows improvement compared to linear and polynomial, with eight stunting cases detected and some detection errors in normal cases. Overall, the RBF and sigmoid kernels are more capable of capturing stunting cases compared to the linear and polynomial kernels, making them more effective for stunting data classification in this study.

2) SVM Implementation with Balanced Parameter

To address class imbalance in the dataset, SVM was also applied with class weights balanced using the parameter 'class weight=balanced.' The process followed the same steps as in the previous phase. The evaluation results from this implementation are presented in Fig. 7. The application of class weights showed changes in the performance of minority class detection, with varying impacts depending on the kernel type used.

With the application of the parameter 'class weight = balanced,' there is a significant change in the performance of minority class detection, as shown in Fig. 7. The Linear kernel shows improvement in stunting detection, with 26 stunting cases detected compared to 0 previously, although there is a decrease in normal case detection. The

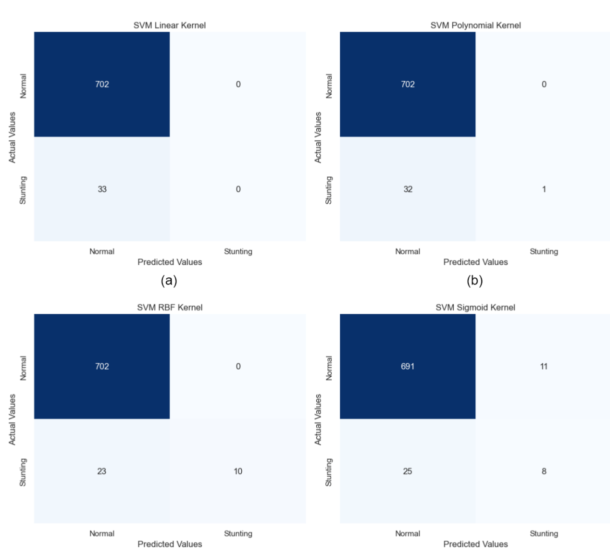

Fig. 6. SVM without balanced parameter confusion matrix: (a) linear, (b) polynomial, (c) RBF, and (d) sigmoid kernels.

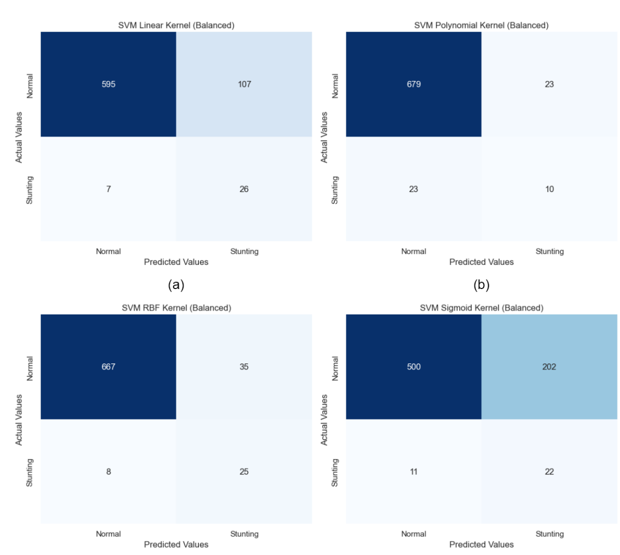

Fig. 7. SVM with balanced parameter confusion matrix: (a) linear, (b) polynomial, (c) RBF, and (d) sigmoid kernels.

polynomial kernel demonstrates relatively good results with 10 stunting cases detected and a smaller decrease in normal case detection compared to linear. The RBF kernel shows even better performance with 25 stunting cases detected and a moderate decrease in normal case detection. The sigmoid kernel, while detecting 22 stunting cases, experiences a drastic decline in normal case accuracy with 202 cases classified as stunting. Overall, applying class weight balancing improves stunting detection capability across all kernel types, with the RBF kernel showing the best balance between stunting and normal detection.

3) Decision Tree Implementation

In addition to SVM, the decision tree algorithm was implemented to classify stunting data using the scikit-learn library. The model was trained and evaluated, with results presented in Fig. 8. The Decision Tree demonstrated its ability to handle class imbalance, particularly in identifying the 'Normal' and 'Stunting' classes with satisfactory results.

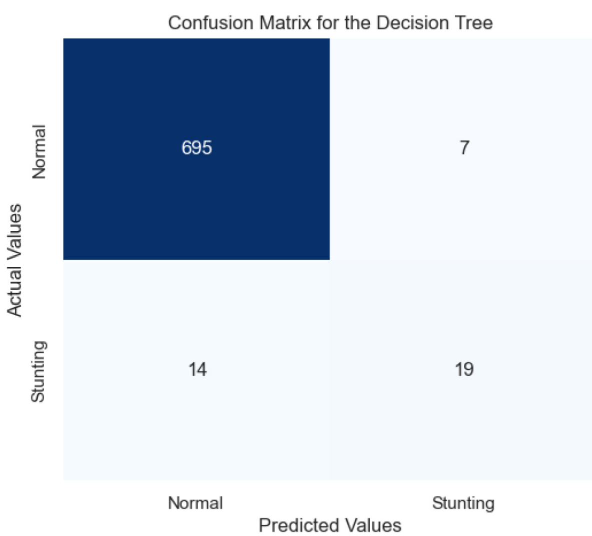

Fig. 8. Decision tree confusion matrix.

The implementation of the decision tree algorithm demonstrates its capability in handling class imbalance quite effectively. Based on the confusion matrix presented in Fig. 8, the decision tree successfully detected 19 stunting cases and 695 normal cases. Although there were some classification errors with 14 stunting cases incorrectly classified as normal, these results still indicate relatively good performance in identifying both classes. The decision tree shows adequate ability in addressing class imbalance without the need for special parameters, yielding better results compared to several SVM kernels in terms of minority class detection.

4) Non-SMOTE Accuracy

The classification results Table 1 on the imbalanced dataset show that while the SVM model with the RBF kernel achieved the highest accuracy of 96.87%, the F1-score for the stunting class was only 47%. This indicates that despite the high overall accuracy, the RBF SVM model was not effective in detecting stunting cases. In comparison, the Decision Tree without SMOTE demonstrated an accuracy of 97.14% with an F1-score of 64%, which represents a better result but still shows that stunting detection was not optimal.

C. SMOTE Implementation

1) SVM SMOTE Implementation

To handle class imbalance in the stunting dataset, SVM was implemented with the SMOTE balancing technique. This technique aims to improve class distribution by creating synthetic samples of the minority class, resulting in a more balanced dataset. The SVM implementation using SMOTE followed the same procedures as SVM without SMOTE, with the dataset first processed to address class imbalance.

The application of the SMOTE technique in SVM, as shown in Fig. 9, significantly improved the detection

Table 1. Non-SMOTE model performance metrics (unit: %)

Model	Kernel	Class weight	Accuracy	Precision (stunting)	Recall (stunting)	F1-score (stunting)
SVM	Linear	None	95.51	0	0	0
	Polynomial	None	95.65	100	3.00	6.00
	RBF	None	96.87	100	30.00	47.00
	Sigmoid	None	95.10	42.00	24.00	31.00
	Linear	Balanced	84.49	20.00	79.00	31.00
	Polynomial	Balanced	93.74	30.00	30.00	30.00
	RBF	Balanced	94.15	42.00	76.00	54.00
	Sigmoid	Balanced	71.02	10.00	67.00	17.00
Decision tree	-	None	97.14	73.00	58.00	64.00

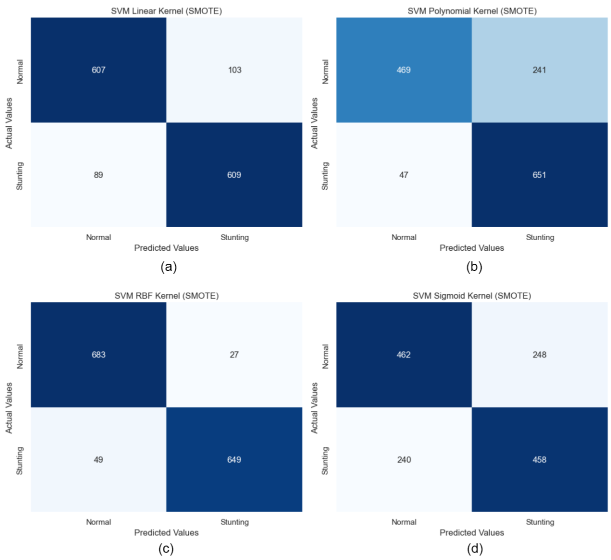

Fig. 9. SVM SMOTE confusion matrix: (a) linear, (b) polynomial, (c) RBF, and (d) sigmoid kernels.

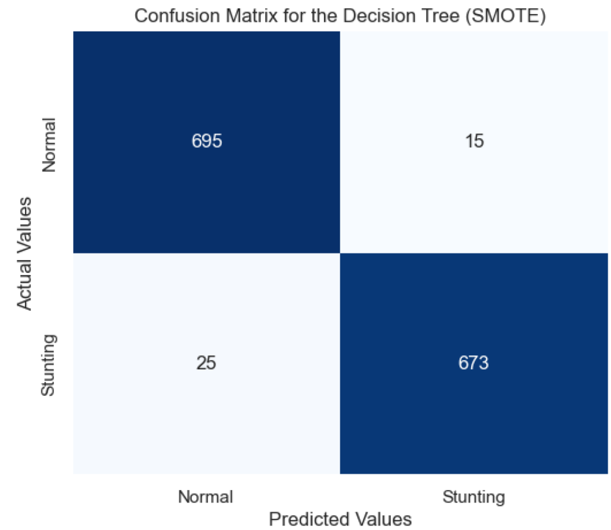

Fig. 10. Decision tree SMOTE confusion matrix.

capability for stunting across all kernel types, with varying results among kernels. The linear kernel showed an increase in stunting detection to 609 cases, but with an increase in normal detection errors. The polynomial kernel also demonstrated improvement with 651 stunting cases detected, although at the cost of normal accuracy. The RBF kernel provided the best results, detecting 649 stunting cases with minimal normal detection errors, while the sigmoid kernel, despite enhancing stunting detection, experienced a substantial decrease in normal detection accuracy. Overall, SMOTE enhanced stunting detection performance, with the RBF kernel showing the best balance between sensitivity and specificity.

2) Decision Tree SMOTE Implementation

To enhance the performance of the decision tree in dealing with class imbalance, the SMOTE technique was applied to the dataset before model training. The implementation of the decision tree with SMOTE followed the same methods as the decision tree without SMOTE, with an additional step of data balancing using SMOTE to ensure a more balanced class distribution before training the model.

The application of the SMOTE technique in the decision tree algorithm showed a significant improvement in detecting the minority class of stunting. Based on Fig. 10, the decision tree with SMOTE successfully detected 673 stunting cases and 695 normal cases, with the number of stunting detection errors reduced to 25. The SMOTE technique improved class balance in the dataset, allowing the decision tree to identify stunting more effectively without sacrificing much accuracy in normal detection. These results indicate that SMOTE is effective in enhancing the performance of the decision tree in handling class imbalance in the stunting dataset.

3) SMOTE Accuracy

When SMOTE was applied to the dataset, the classification results showed a significant improvement in performance. Table 2 illustrates that after applying SMOTE, the SVM model with the RBF kernel achieved an accuracy of 94.60% and an F1-score of 94%. This indicates that SMOTE enhanced the model's ability to detect stunting cases more accurately, with a higher F1-score compared to when the data was imbalanced. Meanwhile, the decision tree with SMOTE achieved the most impressive results with an accuracy of 96.95% and an F1-score of 97%.

Table 2. SMOTE model performance metrics (unit: %)

Model	Kernel	Accuracy	Precision (stunting)	Recall (stunting)	F1-score (stunting)
SVM	Linear	86.36	86.00	87.00	86.00
	Polynomial	79.55	73.00	93.00	82.00
	RBF	94.60	96.00	93.00	94.00
	Sigmoid	65.34	65.00	66.00	65.00
Decision tree	-	96.95	98.00	96.00	97.00

These results show that the decision tree, when combined with SMOTE, not only maintained high accuracy but also significantly improved performance in detecting stunting.

D. Accuracy Comparison

In this study, the performance of classification methods for the stunting dataset was evaluated both with and without the application of the SMOTE technique for data balancing.

This comparison indicates that the application of SMOTE effectively addressed the data imbalance issue and improved the overall classification performance Table 3. The decision tree with SMOTE proved to be the most effective method for stunting classification, offering a combination of high accuracy and good detection capability. These results suggest that using data balancing techniques like SMOTE can significantly enhance classification outcomes on imbalanced datasets, especially in the context of detecting stunting cases.

E. Testing

In this study, performance and accuracy of the developed models were tested on original data with both stunting and non-stunting labels. The testing aimed to evaluate the model's accuracy and effectiveness in detecting stunting among infants using samples from Bandarharjo Community Health Centers. The goal was to ensure that the models could accurately classify stunting and non-stunting cases according to their labels.

1) Stunting Prediction Testing

The stunting prediction results were tested using five sample data points that include important variables such as age, weight, and height, as shown in Table 4.

Table 3. Best performance metrics by method (unit: %)

Method	Accuracy	F1-score (macro avg)
SVM - RBF	96.87	72.00
SVM - RBF (balanced)	94.15	75.00
SVM - RBF (SMOTE)	94.60	95.00
Decision tree	97.14	81.00
Decision tree (SMOTE)	97.16	97.00

Table 4. Sample stunting dataset for prediction testing

	Age (mo)	Weight (kg)	Height (cm)
Children 1	5	6.7	62.0
Children 2	4	6.4	59.0
Children 3	28	9.2	80.6
Children 4	13	9.0	70.3
Children 5	47	10.1	88.0

Table 5. Stunting prediction results by method

Model	Data 1	Data 2	Data 3	Data 4	Data 5
NON-SMOTE					
SVM Linear	Normal	Normal	Normal	Normal	Normal
SVM Polynomial	Normal	Normal	Normal	Normal	Normal
SVM RBF	Normal	Normal	Stunting	Normal	Stunting
SVM Sigmoid	Normal	Normal	Normal	Normal	Normal
SVM Linear (balanced)	Stunting	Stunting	Stunting	Normal	Stunting
SVM Polynomial (balanced)	Normal	Normal	Normal	Normal	Stunting
SVM RBF (balanced)	Normal	Normal	Stunting	Normal	Stunting
SVM Sigmoid (balanced)	Normal	Normal	Stunting	Normal	Stunting
Decision tree	Normal	Stunting	Stunting	Stunting	Stunting
SMOTE					
SVM Linear (SMOTE)	Normal	Stunting	Stunting	Normal	Stunting
SVM Polynomial (SMOTE)	Normal	Normal	Stunting	Normal	Stunting
SVM RBF (SMOTE)	Normal	Normal	Stunting	Normal	Stunting
SVM Sigmoid (SMOTE)	Normal	Normal	Stunting	Normal	Stunting
Decision tree (SMOTE)	Normal	Normal	Stunting	Stunting	Stunting

The results of stunting prediction for these data points are shown in Table 5. Table 5 presents the stunting prediction results using various methods, both with data balancing techniques (SMOTE) and without.

From Table 5, it can be seen that the decision tree method shows strong performance in detecting stunting cases, with consistent predictions of stunting for most data points, particularly in both non-SMOTE and SMOTE scenarios. However, the predictions for the second data point in the SMOTE scenario and the fourth data point in the non-SMOTE scenario differ, as these points were classified as normal.

The SVM methods show varied results depending on the kernel type used and whether the SMOTE technique was applied.

- SVM with linear kernel: Most data points are classified as normal without SMOTE, but with SMOTE, only the first and fourth data points are predicted as normal.
- SVM with polynomial kernel: Tends to classify most data points as normal, both with and without SMOTE. Only the third and fifth data points are classified as stunting after applying SMOTE.
- SVM with RBF kernel: Shows inconsistency, with some data points classified as stunting, particularly the third and fifth, both with and without SMOTE.
- SVM with sigmoid kernel: Classifies most data as normal without SMOTE; however, with SMOTE, some data points are classified as stunting, particularly the third and fifth.

This analysis shows that the decision tree method is more stable and consistent in stunting classification compared to various SVM variants, especially without SMOTE application. Meanwhile, the SMOTE technique has varying effects on the performance of SVM models, highlighting the importance of selecting appropriate methods and data balancing techniques in the model testing and evaluation process. These results indicate that the decision to use SMOTE should be based on specific evaluations of the methods used and the type of data encountered.

2) Normal Prediction Testing

For the normal prediction test, five sample data points were used, consisting of information on age, weight, and height, as shown in Table 6. These sample data reflect normal conditions based on the available parameters and aim to test the model's accuracy in classifying normal data.

Table 6. Sample normal dataset for prediction testing

	Age (mo)	Weight (kg)	Height (cm)
Child 1	11	8.4	72.0
Child 2	10	9.2	71.3
Child 3	16	9.2	78.3
Child 4	26	11.1	87.9
Child 5	40	13.6	95.8

Table 7. Normal prediction results by method

Model	Data 1	Data 2	Data 3	Data 4	Data 5
NON-SMOTE					
SVM Linear	Normal	Normal	Normal	Normal	Normal
SVM Polynomial	Normal	Normal	Normal	Normal	Normal
SVM RBF	Normal	Normal	Normal	Normal	Normal
SVM Sigmoid	Normal	Normal	Normal	Normal	Normal
SVM Linear (balanced)	Normal	Normal	Normal	Normal	Normal
SVM Polynomial (balanced)	Normal	Normal	Normal	Normal	Normal
SVM RBF (balanced)	Normal	Normal	Normal	Normal	Normal
SVM Sigmoid (balanced)	Normal	Normal	Normal	Normal	Normal
Decision tree	Normal	Normal	Normal	Normal	Normal
SMOTE					
SVM Linear (SMOTE)	Normal	Normal	Normal	Normal	Normal
SVM Polynomial (SMOTE)	Normal	Normal	Normal	Normal	Normal
SVM RBF (SMOTE)	Normal	Normal	Normal	Normal	Normal
SVM Sigmoid (SMOTE)	Normal	Normal	Normal	Normal	Normal
Decision tree (SMOTE)	Normal	Normal	Normal	Normal	Normal

The prediction results for this normal data can be seen in Table 7, which shows the outcomes from various methods, both with and without the application of SMOTE techniques.

From Table 7, it can be seen that most methods correctly classify normal data, particularly the SVM linear, SVM Polynomial, and SVM RBF methods, both with and without the application of SMOTE.

The decision tree method also shows consistent performance, predicting normal for all samples in both SMOTE and non-SMOTE scenarios.

However, there are some inconsistencies in the SVM with sigmoid kernel (balanced) and SVM sigmoid (SMOTE) methods where some data points are classified as stunting. In SVM sigmoid (balanced), the third data point is classified as stunting, while in SVM sigmoid (SMOTE), the third and fourth data points are classified as stunting. This indicates that SVM with sigmoid kernel may be less suitable for detecting normal data in this testing context.

Additionally, the SVM polynomial (SMOTE) method shows inconsistency with the fifth data point being classified as stunting. This highlights the importance of selecting the appropriate kernel to achieve optimal results, especially when data balancing techniques such as SMOTE are used.

Overall, this testing demonstrates that some methods are more stable and accurate in detecting normal cases, particularly the decision tree method and SVM variants with linear and RBF kernels. The SMOTE technique does not appear to significantly affect the results for most methods, except for some SVM variants using sigmoid and polynomial kernels.

This normal prediction test is essential to ensure that the model is not only focused on stunting detection but also maintains good performance in detecting normal conditions. This is crucial to ensure that the model can be reliably used in various data conditions, both for stunting detection and for identifying normality in toddlers.

IV. CONCLUSION

This study evaluates the performance of SVM and decision tree models in detecting stunting among toddlers using a dataset from the Bandarharjo Community Health Centers. The analysis was conducted on both the imbalanced dataset and after applying the SMOTE technique for data balancing.

In the initial stage with the imbalanced data, the SVM model with the RBF kernel showed a relatively high accuracy of 96.87%, but it struggled with stunting class detection, reflected in a low F1-score of 47%. Other SVM models, including those with linear, polynomial, and sigmoid kernels, also showed limited ability to improve stunting detection, even with adjusted class weights. The decision tree without SMOTE showed slightly better performance with an accuracy of 97.14% and an F1-score of 64%, indicating it was more effective in detecting stunting than the SVM models in the imbalanced scenario, but still far from optimal.

After applying SMOTE to address data imbalance, there were notable improvements in the models' abilities to detect stunting. The SVM with the RBF kernel improved to an accuracy of 94.60% and an F1-score of 94%, showing better stunting detection. The decision tree model with SMOTE outperformed all, with an accuracy of 96.95% and an F1-score of 97%. However, while the detection of the minority class improved, some prediction errors persisted, particularly in the normal class.

Further testing with both stunting and normal prediction data revealed that the decision tree method was more stable and consistent in classifying stunting and normal cases compared to SVM variants, especially without the SMOTE application. The SVM models with linear and RBF kernels showed relatively consistent results for normal data, while SVM with sigmoid kernel displayed inconsistencies, misclassifying some normal data points as stunting.

In conclusion, while the application of SMOTE significantly improved the sensitivity of models to stunting cases, the performance enhancement was not uniform across all scenarios. Certain SVM variants, particularly those with sigmoid and polynomial kernels, exhibited prediction inconsistencies even after data balancing. This suggests that while SMOTE is valuable for addressing data imbalance, it does not guarantee comprehensive performance improvements in all prediction aspects. Therefore, additional measures such as model adjustments, parameter tuning, or exploring alternative data balancing techniques may be necessary to further reduce prediction errors and enhance overall model reliability in diverse prediction conditions.

CONFLICT OF INTEREST

The authors have declared that no competing interests exist.

REFERENCES

1. C. R. Titaley, I. Ariawan, D. Hapsari, A. Muasyaroh, and M. J. Dibley, "Determinants of the stunting of children under two years old in Indonesia: a multilevel analysis of the 2013 Indonesia basic health survey," *Nutrients*, vol. 11, no. 5, article no. 1106, 2019. https://doi.org/10.3390/nu11051106
2. J. Mayneris-Perxachs and J. R. Swann, "Metabolic phenotyping of malnutrition during the first 1000 days of life," *European Journal of Nutrition*, vol. 58, no. 3, pp. 909-930, 2019. https://doi.org/10.1007/s00394-018-1679-0
3. J. Aurima, S. Susaldi, N. Agustina, A. Masturoh, R. Rahmawati, and M. T. M. Madhe, "Faktor-faktor yang berhubungan

dengan kejadian stunting pada balita di Indonesia [Factors associated with stunting in toddlers in Indonesia]," *Open Access Jakarta Journal of Health Sciences*, vol. 1, no. 2, pp. 43-48, 2021. https://doi.org/10.53801/oajjhs.v1i3.23

4. M. G. Daffa and P. H. Gunawan, "Stunting classification analysis for toddlers in Bojongsoang: a data-driven approach," in *Proceedings of 2024 2nd International Conference on Software Engineering and Information Technology (ICoSEIT)*, Bandung, Indonesia, 2024, pp. 42-46. https://doi.org/10.1109/ICoSEIT60086.2024.10497515
5. P. Webb, H. Luo, and U. Gentilini, "Measuring multiple facets of malnutrition simultaneously: the missing link in setting nutrition targets and policymaking," *Food Security*, vol. 7, pp. 479-492, 2015. https://doi.org/10.1007/s12571-015-0450-0
6. A. T. A. Sibuea and P. H. Gunawan, "Classifying stunting status in toddlers using k-nearest neighbor and logistic regression analysis," in *Proceedings of 2024 International Conference on Data Science and Its Applications (ICoDSA)*, Kuta, Bali, Indonesia, 2024, pp. 6-11. https://doi.org/10.1109/ICoDSA62899.2024.10652063
7. G. A. F. Khansa and P. H. Gunawan, "Predicting stunting in toddlers using KNN and naïve Bayes methods," in *Proceedings of 2024 International Conference on Data Science and Its Applications (ICoDSA)*, Kuta, Bali, Indonesia, 2024, pp. 17-21. https://doi.org/10.1109/ICoDSA62899.2024.10651676
8. H. Janawisuta and P. H. Gunawan, "Early detection of stunting in Indonesian toddlers: a machine learning approach," in *Proceedings of 2024 International Conference on Data Science and Its Applications (ICoDSA)*, Kuta, Bali, Indonesia, 2024, pp. 12-16. https://doi.org/10.1109/ICoDSA62899.2024.10651637
9. J. R. Khan, J. H. Tomal, and E. Raheem, "Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh," *Informatics for Health and Social Care*, vol. 46, no. 4, pp. 425-442, 2021. https://doi.org/10.1080/17538157.2021.1904938
10. N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak [Analysis of performance comparison of machine learning models for predicting stunting risk in children's growth]," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 413-422, 2024.
11. D. Septhya, K. Rahayu, S. Rabbani, V. Fitria, R. Rahmaddeni, Y. Irawan, and R. Hayami, "Implementasi algoritma decision tree dan support vector machine untuk klasifikasi penyakit kanker paru [Implementation of decision tree algorithm and support vector machine for lung cancer classification]," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 15-19, 2023.
12. H. Tanveer, M. A. Adam, M. A. Khan, M. A. Ali, and A. Shakoor, "Analyzing the performance and efficiency of machine learning algorithms, such as deep learning, decision trees, or support vector machines, on various datasets and applications," *The Asian Bulletin of Big Data Management*, vol. 3, no. 2, pp. 126-136, 2023. https://doi.org/10.62019/abbdm.v3i2.83
13. S. Amarappa and S. V. Sathyanarayana, "Data classification using support vector machine (SVM), a simplified approach," *International Journal of Electronics and Computer Science Engineering*, vol. 3, pp. 435-445, 2014.
14. T. Vaivada, N. Akseer, S. Akseer, A. Somaskandan, M. Stefopulos, and Z. A. Bhutta, "Stunting in childhood: an overview of global burden, trends, determinants, and drivers of decline," *The American Journal of Clinical Nutrition*, vol. 112, pp. 777S-791S, 2020. https://doi.org/10.1093/ajcn/nqaa159
15. I. P. Putri, T. Terttiaavini, and N. Arminarahmah, "Analisis perbandingan algoritma machine learning untuk prediksi stunting pada anak [Comparative analysis of machine learning algorithms for predicting child stunting]," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 257-265, 2024.
16. A. Y. Perdana, R. Latuconsina, and A. Dinimaharawati, "Prediksi Stunting Pada Balita Dengan Algoritma Random Forest [Predicting stunting in toddlers using the random forest algorithm]," *eProceedings of Engineering*, vol. 8, no. 5, pp. 6650-6656, 2021. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16455
17. P. Soltanzadeh and M. Hashemzadeh, "RCSMOTE: range-controlled synthetic minority over-sampling technique for handling the class imbalance problem," *Information Sciences*, vol. 542, pp. 92-111, 2021. https://doi.org/10.1016/j.ins.2020.07.014
18. M. Mukherjee and M. Khushi, "SMOTE-ENC: a novel smote-based method to generate synthetic data for nominal and continuous features," *Applied System Innovation*, vol. 4, no. 1, article no. 18, 2021. https://doi.org/10.3390/asi4010018
19. D. Jahed Armaghani, Y. Y. Ming, A. Salih Mohammed, E. Momeni, and H. Maizir, "Effect of SVM kernel functions on bearing capacity assessment of deep foundations," *Journal of Soft Computing in Civil Engineering*, vol. 7, no. 3, pp. 111-128, 2023. https://doi.org/10.22115/scce.2023.356959.1510

Aqeela Fathya Najwa https://orcid.org/0009-0003-0265-9597

Aqeela Fathya Najwa is an undergraduate student at the Department of Data Science, School of Computing, Telkom University, Bandung, Indonesia. She is currently pursuing a Bachelor's degree in Data Science, and she also works as a teaching assistant related to mathematics and data science at the School of Computing, Telkom University. She actively participates in data analysis projects and internships, focusing on machine learning, business intelligence, and data visualization. Her research interests include machine learning algorithms, big data analysis, and the application of data science in business decision-making.

Putu Harry Gunawan https://orcid.org/0000-0002-3635-894X

Putu Harry Gunawan is an associate professor at the Department of Informatics, School of Computing, Telkom University, Bandung, Indonesia. He received the B.S. in Computational Mathematics from Udayana University, Indonesia in 2009, the M.Sc. dual degree in Computational Science in 2011 from ITB, Indonesia and Kanazawa University, Japan, and the Doctoral dual degree in Applied Mathematics from ITB and Pari-Est, France in 2015. He has authored five local lecture books in numerical mathematics and computational statistics, several publications in top journals, such as *Computers & Fluids*, *Computational Geosciences*, and the *Progress in Computational Fluid Dynamics* (CFD). Moreover, he has authored more than 70 papers for international conferences in IEEE, Springer, Elsevier, and *Journal of Physics: Conference Series*. His research interests are in the field of computational science, numerical methods, modeling and simulation, and parallel scientific computing.