Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

Performance Optimization of Support Vector Machine with SMOTE for Multiclass Stunting Prediction in Sumedang District, Indonesia

Irfan Fadil*[1], Ramdani Surya Manggala[2], Esa Firmansyah[3], Muhammad Agreindra Helmiawan[4]

[1,2]Informatics, Universitas Sebelas April, Indonesia
[3,4]Information and Communication Technology, School of Science and Technology Asia e University, Subang Jaya, Malaysia

Email: [1]fadilirfan@unsap.ac.id

Received : Jun 8, 2025; Revised : Sep 23, 2025; Accepted : Sep 25, 2025; Published : Sep 27, 2025

Abstract

The percentage of stunting toddlers in Sumedang Regency is the highest compared to other nutritional problems. Stunting imposes a significant risk to the future quality of human resources. This study explores the performance of the Support Vector Machine (SVM) algorithm in predicting the stunting status of toddlers in Tanjungmedar Subdistrict, the region with the highest incidence of stunting cases in Sumedang Regency in 2020. The testing uses RapidMiner software and applies the Synthetic Minority Oversampling Technique (SMOTE) to overcome the imbalanced dataset so that the resulting performance can be optimized. Accuracy, precision, recall, and F1-score are measured in performance evaluation using a confusion matrix. The findings demonstrate that SMOTE might adjust the distribution of target classes in the dataset to maximize the SVM algorithm's performance. At the start of the test, the SVM model produced an accuracy of 85.10%. After applying SMOTE, the accuracy of the SVM model increased to 89.08%. The F1-score also increased for each class, except for the Normal class, which decreased slightly. These results demonstrate the suitability of SVM combined with SMOTE for health-related multiclass classification tasks, especially in imbalanced public health datasets, contributing to the advancement of applied machine learning in healthcare informatics.

Keywords : *Classifying, Confusion Matrix, RapidMiner, Stunting Treatment.*

1. INTRODUCTION

The COVID-19 pandemic in Indonesia has had a profound impact on multiple sectors, including the economy, education, and various aspects of social life, such as health-related concerns [1], [2], [3]. The current state of health facilities is characterized by significant overload, while disruptions in food supply chains persist. Furthermore, the economic repercussions of COVID-19 may lead to a notable increase in the incidence of nutritional issues among children in Indonesia. Even before the COVID-19 pandemic, Indonesia was already facing high levels of malnutrition [4].

Indonesia is one of the countries with diverse nutritional problems. Currently, Indonesia is still working hard to overcome nutritional issues, one of which is stunting or short stature [5]. Stunting represents a developmental impairment in infants under five years of age, commonly referred to as toddlers, arising from prolonged malnutrition. This condition manifests as a height or length that is insufficient relative to their chronological age [6]. Many people are unaware that short stature is a sign of chronic malnutrition in children [7].

Stunting is not only a problem of physical growth disorders, but also causes children to be more susceptible to disease, have suboptimal intelligence, and in the future may be at risk of decreased productivity [8]. Stunting can ultimately slow down economic growth, exacerbate poverty, and perpetuate injustice. Thus, stunting is a significant threat to the quality of human resources in Indonesia.

The prevalence of stunting in Indonesia ranks 115th out of 151 countries worldwide, based on stunting data from the World Bank, JME, and UNICEF in 2020. According to the results of the 2021 Indonesia Nutrition Status Study (SSGI), the national stunting rate was 27.7% in 2019, decreasing by 1.6% annually to 24.4% in 2021. However, this figure remains high compared to the WHO threshold of 20% [9].

In Sumedang District, based on the results of the Under-Five Weighing Month (BPB) activities during interviews with the Health Office, it was explained that in 2020, the percentage of stunted children was the highest at 12.05% compared to other nutritional problems, with 7.71% underweight, 3.46% wasting, and 1.34% overweight. The most prevalent number of stunting instances among infants in Sumedang Regency was found in Cibugel District and Tanjungmedar District. In terms of proportion, the severity of infant malnutrition across all categories in Sumedang Regency is below the WHO threshold for malnutrition. However, compared to the 2019 BPB results, there was an increase in the percentage of all categories of infant malnutrition. Given the high prevalence of stunting among infants both nationally and locally in Sumedang District, as well as its adverse effects on future generations, it is crucial to conduct research that can assist in efforts to reduce stunting prevalence. One approach is to predict the stunting status of infants. By identifying infants with stunting, they can be promptly addressed to prevent their condition from worsening and the growth impairments they experience from becoming more severe.

The Research related to predicting stunting status can be conducted using machine learning approaches [10], [11] and data mining approaches using the Random Forest algorithm [5], [12], Artificial Neural Networks [13], Support Vector Machine [14], [15], K-Nearest Neighbor and Forward Chaining [16], [17], Decision Tree C4.5 [18], Combining Naive Bayes Method and the C4.5 Algorithm [19], and Convolutional Neural Network [20]. Several studies have been conducted using different sample data and from other locations. In previous studies, multiclass prediction using a combination of SVM and SMOTE has not been performed. In addition, local datasets from Tanjung Medar have not been used in this earlier study.

The application of Support Vector Machine (SVM) in public health applications is increasingly being discussed. One example of the practical application of SVM is in mental health monitoring, as described by [21]. In addition, SVM is also used in the context of infectious diseases, as shown by [22], where SVM plays a vital role in predicting contagious diseases. The implementation of SVM in innovative health monitoring systems is also highlighted in research by [23]. This study explains how SVM is integrated into an innovative health monitoring system, achieving high accuracy in health predictions. Furthermore, research by [24] highlights the use of machine learning techniques, including SVM, in remote health monitoring in clinical trials. These algorithms not only help identify patterns from transformed data but can also contribute to better clinical decision-making, reducing patient waiting times in hospitals and potentially lowering mortality rates.

In addition, the Support Vector Machine (SVM) algorithm in public health has proven to be effective in various applications, including sentiment analysis of health applications such as JKN Mobile, as demonstrated by [25], as well as mapping public opinion on BPJS Kesehatan policies analyzed by [26]. This study shows that SVM is capable of classifying user reviews with high accuracy, as revealed by [27]. In addition, SVM is also used to analyze sentiment about the COVID-19 vaccine, enabling a deeper understanding of public opinion, as shown by [28] in the context of risk prediction, and research by [29] utilizing web-based applications to identify possible diabetes. By utilizing SVM, various studies have demonstrated its ability to automate data classification and provide valuable insights for service improvement and evidence-based decision-making in public health systems.

The purpose of this study is to explore the performance of the Support Vector Machine (SVM) algorithm in predicting stunting status in toddlers in Tanjungmedar District, the area with the highest

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

incidence of stunting in Sumedang Regency in 2020. This study also aims to address the problem of data imbalance by utilizing the Synthetic Minority Oversampling Technique (SMOTE) to enhance the performance of the SVM model in multi-class health classification. The evaluation is conducted using accuracy, precision, recall, and F1-score metrics based on a confusion matrix.

2. METHOD

In this study, the research methodology employed to make the model predict stunting at the Tanjungmedar Community Health Center, located on Jalan Raya Cikaramas, Tanjungmedar District, Sumedang Regency. The methodology encompasses five key stages: data collection, data selection, preprocessing, data mining, and evaluation. Each stage is systematically structured to address the challenges of handling toddlers' stunting and to optimize the accuracy and interpretability of the predictive models. The following subsections detail the procedures and techniques applied in each phase to achieve the research objectives.

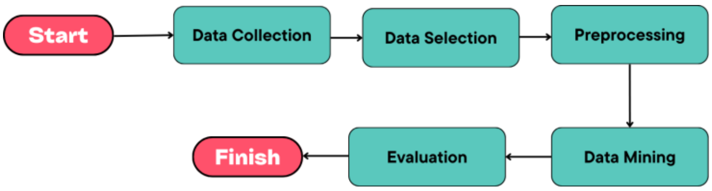

Figure 1. Research Model Design

2.1. Data Collection

The data source for this study was the Tanjungmedar Community Health Center, located on Jalan Raya Cikaramas, Tanjungmedar District, Sumedang Regency. Data was obtained through the Community-Based Nutrition Recording and Reporting (e-PPGBM) application, which is part of the Indonesian Ministry of Health's Integrated Nutrition Information System (called "Sigizi Terpadu"). In Figure 2, there is a screenshot of the e-PPGBM application.

The data collected is nutritional status data for toddlers from 2020 to 2021 from three villages with the highest cases of stunting, namely Kamal Village, Tanjungmedar Village, and Tanjungwangi Village. The data was collected using a nutritional status filter for TB/U with the following values: Stunting (Very Short and Short), Normal, and High. The following is a screenshot of the menu from the e-PPGBM application used to collect the required data.

Figure 2. Screenshot e-PPGBM Application

2.2. Data Selection

At this stage, the necessary data has been collected, and the attributes used for the data mining process have been determined. Redundant attributes will be eliminated at this stage. Redundant attributes or irrelevant attributes are attributes that are not important to consider because they do not correlate with the target class attributes so that they can be eliminated in the categorization/classification.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

2.3. Preprocessing

At this stage, preprocessing is performed on the obtained data, with the following objectives:

- To make the data easier to understand.
- To improve data quality so that the results of data mining are better.
- To improve the efficiency and ease of the data mining process.
- The dataset can be used for the data mining process.

The techniques used in this study are as follows.

- Data transformation

Raw data has attribute values that the model cannot process. Therefore, attributes containing values that the SVM algorithm cannot process will be changed/converted into values that the algorithm can process.

- Synthetic Minority Oversampling Technique (SMOTE)

With the SMOTE technique, the minority class in the dataset is "multiplied" so that the class distribution becomes balanced.

- Cleaning outliers

A dataset can be considered dirty if it contains outliers. Generally, data mining results are poor when using a dirty dataset. Therefore, before the data mining process, outliers in the dataset must be removed first.

2.4. Data Mining

Once the preprocessing procedure has been completed, the "clean" dataset is used in this stage. In the data mining stage, modeling and testing of the model/classifier from the SVM algorithm are carried out using RapidMiner software. Testing is conducted using two scenarios: SVM without SMOTE application and SVM with SMOTE application, to compare the results. Thus, the impact of SMOTE application on model performance can be determined.

2.5. Evaluation

After conducting data mining, or in other words, testing the SVM algorithm, the performance values obtained are interpreted at this stage. Based on the performance obtained, the SVM algorithm can be evaluated for its suitability in predicting the status of stunting in toddlers.

3. RESULT

3.1. Data Collection

The amount of data obtained during the data collection stage was 2,044 data points. Of these, 600 data points were for stunted toddlers, 1,438 data points were for typical toddlers, and 6 data points were for tall toddlers. Figure 3 illustrates the percentage of data.

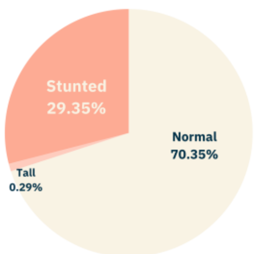

Figure 3. Percentage of nutritional status data for infants based on Height by Age (called TB/U) in 2020-2021

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

The collected data contains 33 attributes, namely: Number, Population Identification Number, Name, Gender, Date of Birth, Birth Weight, Birth Height, Parent's Name, Complete Address Consisting of: Province, Regency/City, Sub-district, Health Center, Village/Sub-district, Integrated Health Post, Age, Measurement Date, Weight, Height, LiLA, BB/A, ZS BB/A, TB/A, TB/A, BB/TB, ZS BB/TB, Weight Gain, PMT Received (kg), Amount of Vit A, KPSP, and KIA.

3.2. Data Selection

Attributes that are considered redundant are not used as source data. After separation, the attribute data used for the data mining process are Gender, Age at Measurement (called Age), Height, Z-Score Height by Age (TB/U) (called Z-Score), and Height by Age (TB/U). The structure of the dataset is presented in Table 1.

Table 1. Dataset After Data Selection

Gender	Age	Height	Z-Score	TB/U
Female	4 Year - 9 Month - 3 Day	95.8	-2.6	Stunted
Female	4 Year - 9 Month - 26 Day	102	-1.34	Normal
Male	4 Year - 9 Month - 23 Day	102.1	-1.46	Normal
Male	4 Year - 9 Month - 23 Day	92.5	-3.57	Severely Stunted
....	....	....	....	....
Female	4 Year - 9 Month - 23 Day	101.8	-1.38	Normal

3.3. Preprocessing

After determining the attributes used for the data mining process, the next step is preprocessing, which involves cleaning/preparing the values of existing attributes so that they are simpler for the model to comprehend and can enhance the model's performance. The techniques applied are as follows.

- Data transformation

As shown in Table 1, the value of the Age attribute is a description of age in years, months, and days. For the Age data to be processed by the model, it needs to be converted to a more specific unit of measurement. In this study, the Age values were converted to a unit based only on months. Thus, the dataset became as shown in Table 2.

Table 2. Dataset after Age values have been converted

Gender	Age	Height	Z-Score	TB/U
Female	57.1	95.8	-2.6	Stunted
Female	57.86	102	-1.34	Normal
Male	57.76	102.1	-1.46	Normal
Male	57.76	92.5	-3.57	Severely Stunted
....	....	....	....	....
Female	57.76	101.8	-1.38	Normal

- SMOTE

After all attributes have values that the model can process, the next step is to apply the SMOTE technique to handle imbalanced data—the distribution of target classes in the dataset before SMOTE is illustrated in Table 3.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

Table 3. Class distribution before SMOTE

Class Name	Amount of Data
Normal	57.1
Short	57.86
Very Short	57.76
Tall	57.76

To apply the SMOTE technique, first import the dataset into RapidMiner software. Import the dataset with each column/attribute set based on the data type of the attribute value, namely Gender as binomial, Age as integer, Height as real, Z-Score as real, and TB/U as polynomial. Additionally, the attribute that serves as the class/label is TB/U, which has four values: Very Short, Short, Normal, and Tall. It is crucial to remember that the imported dataset contains no missing values. Therefore, the oversampling process with SMOTE can be performed directly.

The application of SMOTE on the dataset utilizes the SMOTE operator available in RapidMiner. Because this operator can only handle one minority class, three SMOTE operators are used to perform oversampling on the tall, very short, and short classes.

Now the distribution of classes is balanced. Each class has 1,438 data points, adjusting to the majority class, which is Normal. Thus, the dataset has a total of 5,752 data points, including synthetic data from SMOTE in the Very Short, Short, and High classes

- Cleaning outliers

To clean the dataset from outliers, two operators were used, as shown in Fig. 4. The two operators added were the Detect Outlier (Distances) operator, which is based on the distance to the k nearest neighbors, and helps discover n outliers. This investigation used the default values of n=10 and k=10 for n and k.

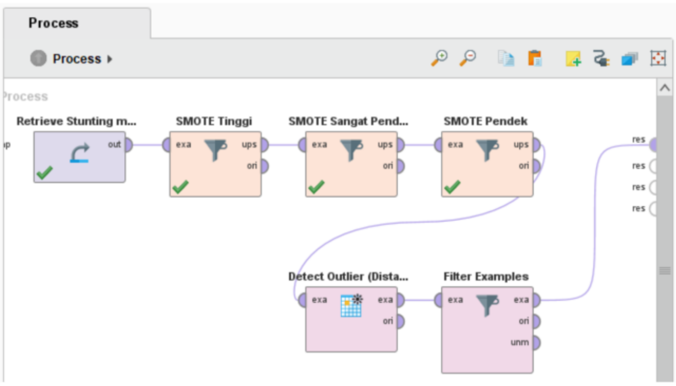

Figure 4. Application of operators to detect outliers

The second operator, Filter Examples, filters the data to be retained or discarded. In this study, the Filter Examples operator was used to maintain data that were not outliers. Thus, the model only uses clean data (without outliers) in the data mining process.

3.4. Data Mining

To test the SVM algorithm, first add the cross-validation operator to the process that has been created. The number of folds used in the cross-validation operator is 10 (k=10). Then, to create a model from the SVM algorithm, double-click on the cross-validation operator. This will bring up a separate process from the cross-validation operator.

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

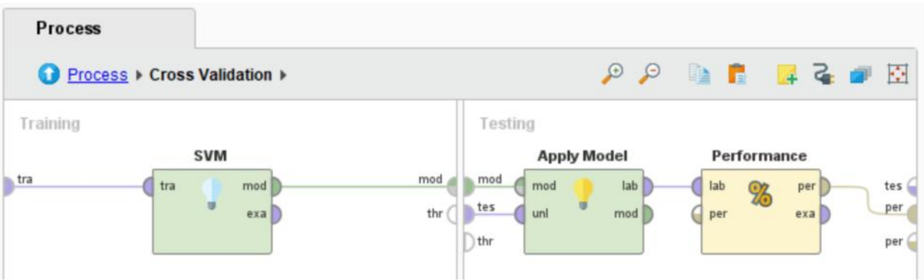

Figure 5. Process design in cross-validation

It can be seen that the cross-validation process consists of two components: Training and Testing. During the training phase, an SVM operator is used, which becomes the model/classifier for this study. Because the SVM algorithm cannot handle attributes with nominal values, another operator needs to be added to the primary process, namely the Nominal to Numerical operator, which functions to convert nominal or categorical data into numerical data. The attribute converted into numerical data is Gender. The design of the primary process after adding the Nominal to Numerical operator is as follows.

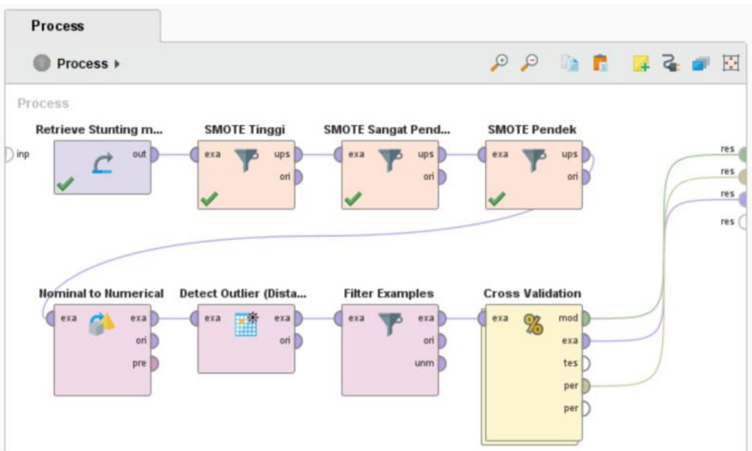

Figure 6. Design process after adding the operator nominal to the numerical

After applying the Nominal to Numerical operator, the Gender attribute becomes two, namely Gender = L and Gender = P, where the value is between 1 (for male or female) and 0 (for neither male nor female). Returning to the cross-validation process, in the Testing section, there is an Apply Model operator that functions to apply the SVM model that has been created. Specifically, it tests the SVM model using the cross-validation method. Finally, the Performance operator evaluates the model's efficacy based on the test outcomes.

3.5. Evaluation

To perform SVM testing without SMOTE, the three operators in the process design must be disabled first. Then, the process was run, and the SVM test results were as follows.

accuracy: 85.10% +/- 1.01% (micro average: 85.10%)

	true Pendek	true Normal	true Sangat Pendek	true Tinggi	class precision
pred. Pendek	301	13	112	0	70.66%
pred. Normal	150	1420	24	4	88.86%
pred. Sangat Pendek	0	0	10	0	100.00%
pred. Tinggi	0	0	0	0	0.00%
class recall	66.74%	99.09%	6.85%	0.00%	

Figure 7. SVM test results

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

To perform testing with SMOTE, the SMOTE operator is reactivated. After the process is run, the results of the SVM + SMOTE testing are as follows.

accuracy: 89.08% +/- 1.55% (micro average: 89.08%)

	true Pendek	true Normal	true Sangat Pendek	true Tinggi	class precision
pred. Pendek	1118	187	59	0	81.96%
pred. Normal	179	1236	48	4	84.25%
pred. Sangat Pendek	141	1	1327	0	90.33%
pred. Tinggi	0	8	0	1434	99.45%
class recall	77.75%	86.31%	92.54%	99.72%	

Figure 8. Results of SVM + SMOTE testing

A comparison table can be constructed based on the outcomes of the two aforementioned test situations.

Table 4. Comparison of test results

	SVM			SVM+SMOTE		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Normal	88.86%	99.09%	0.936966	84.25%	86.31%	0.8526766
Short	70.66%	66.74%	0.686441	81.96%	77.75%	0.797995
Very Short	100%	6.85%	0.128217	90.33%	92.54%	0.914216
Tall	0.00%	0.00%	-	99.45%	99.72%	0.995848
Accuracy		85.10%			89.08%	

The interpretation that can be drawn based on the comparison table above is:

- SVM has very weak values for precision and recall in the High class, and does not even have any values at all. And the F1-score is certainly affected.
- SVM has excellent precision in predicting the Very Short class, but its recall ability in the Very Short class is very poor. As a result, the F1-score is only 0.128217, showing that the Very Short class harmonic mean of precision and recall is extremely low.
- SVM performs well only in predicting the Short and Normal classes.
- Although the increase in Accuracy is not significant, SVM + SMOTE actually has more consistent performance in predicting each class, including Very Short, Short, Normal, and High.
- The Very Short and High classes, which were the shortcomings of SVM, obtained very high F1-score values in the SVM + SMOTE test results, with an F1-score value of 0.914216 for the Very Short class and an F1-score value of 0.995848 for the High class.

4. DISCUSSIONS

This section presents an interpretation of the results obtained and compares them with previous studies. The results of this study provide important insights into the performance of the Support Vector Machine (SVM) algorithm in predicting stunting status in toddlers, especially when faced with the challenge of data imbalance. Optimization using the Synthetic Minority Oversampling Technique (SMOTE) showed a significant improvement in the classification ability of the model. The following discussion will outline the main findings of the study, compare them with existing literature, and explore the implications and limitations of the study.

This study aims to optimize the performance of the Support Vector Machine (SVM) algorithm in predicting stunting status in toddlers, especially by overcoming the challenge of data imbalance using the Synthetic Minority Oversampling Technique (SMOTE). The results obtained show that the application of SMOTE significantly improves the SVM model's ability to classify the nutritional status

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

of toddlers, especially for minority classes such as "Very Short" and "Tall". The overall accuracy of the model increased from 85.10% to 89.08% after the application of SMOTE.

The performance of SVM without SMOTE shows a significant weakness in the “High” and “Very Short” classes, with very low precision and recall values, and even zero in the “High” class. This indicates that the model tends to be biased towards the majority class (“Normal”), a common problem in imbalanced datasets. After SMOTE is applied, the F1-score for the “Very Short” class increases drastically to 0.914216 and for the “High” class to 0.995848, indicating a much better predictive ability for these minority classes. This confirms that SMOTE successfully adjusts the distribution of the target classes in the dataset to maximize the performance of the SVM algorithm.

Several previous studies have also explored the use of machine learning algorithms for predicting stunting, often encountering data imbalance problems. In the study by [30], the SVM algorithm was used to classify the stunting status of toddlers in Indonesia, utilizing a dataset of 6,500 data points. With a "linear" kernel, they achieved 82% accuracy, 80% precision, and 86% recall. Although their accuracy was slightly lower than the results of SVM+SMOTE in this study (89.08%), it is essential to note that the study also showed good performance in stunting classification. This study did not explicitly mention the use of SMOTE, but the increase in accuracy from 65.6% to 81% after hypertuning SVM shows a similar effort to optimize the model.

Furthermore, the study by [31] also highlighted the importance of data balancing techniques, such as SMOTE, to improve the accuracy of ML models in detecting stunting. They compared SVM and Decision Tree, and after applying SMOTE, the Decision Tree achieved an F1-score of 97%, while SVM with an RBF kernel increased significantly to an F1-score of 94%. These results align with our findings, which demonstrate that SMOTE is crucial for addressing the issue of data imbalance in stunting prediction and substantially enhances the performance of the SVM model, particularly for the minority class. Another study by [32] evaluating SVM kernels for stunting classification found that the RBF kernel was the most effective, achieving an accuracy of over 90% with an AUC of 0.926. This indicates that kernel selection also plays an essential role in SVM performance, although the study did not explicitly address data imbalance management.

In a study by [33] comparing XGBoost, Random Forest, SVM, and k-NN for stunting detection, data imbalance was also addressed using SMOTE. Although XGBoost was the best-performing model (accuracy 0.8574, recall 0.8914), this study emphasized that the integration of SMOTE with machine learning models significantly improved recall and ROC-AUC metrics, which are critical in healthcare settings to minimize false negatives. This supports our findings on the importance of SMOTE in enhancing the model's ability to accurately identify stunting cases.

Although the overall accuracy of this study of 89.08% is not always the highest compared to some other studies that may achieve higher accuracy with different algorithms or configurations (e.g., CatBoost Classifier achieved 98.47% in study [34]), the focus of this study on multiclass classification (Very Short, Short, Normal, Tall) and handling imbalanced datasets through SMOTE provide significant contributions. Many studies focus on binary classification (stunting vs. non-stunting), whereas this study examines four nutritional classes, which are more complex and relevant for more targeted interventions. The consistent increase in F1-score across classes after SMOTE, especially in the minority class, indicates that this optimized SVM model is very suitable for use as a comprehensive classification or prediction model for stunting status in toddlers.

5. CONCLUSION

Considering the outcomes of the performance test of the SVM algorithm with SMOTE optimization in predicting the status of stunting in toddlers as described above, it can be concluded that: With the application of SMOTE, although the impact on improving Accuracy is not very significant, the

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

model's ability to predict each class, whether Very Short, Short, Normal, or Tall, is more evenly distributed. The SVM algorithm with SMOTE optimization is suitable for use as a classifier in determining the status of stunting in toddlers. Especially in this study, as a multiclass classifier, which produced an Accuracy value of 89. 08% with an F1-score of 0.914216 for the Very Short class, F1-score of 0.797995 for the Short class, F1-score of 0.852676 for the Normal class, and F1-score of 0.995848 for the High class. This indicates the average harmonic value of Precision and Recall from the SVM model.

This approach contributes explicitly to the development of a more accurate and equitable health information system, enabling the identification of children's health risks and facilitating the optimization of early intervention. The use of SVM with SMOTE is also an example of the application of machine learning algorithms that consider issues of fairness and data bias, which are very important in digital health. Further research can explore more advanced machine learning techniques, such as deep learning or ensemble learning, which combines several models to improve accuracy and generalization capabilities, especially in large and heterogeneous health datasets. Furthermore, additional research can integrate these machine learning models into public health information systems to support data-driven decision-making and the development of more effective nutritional intervention policies.

REFERENCES

[1] N. N. Fadilah, K. Dianta, and A. Pratama, “Impact of Covid-19 on Indonesia’s Education System,” in *International Student Conference on Business, Education, Economics, Accounting, and Management (ISC-BEAM)*, 2023, pp. 724–731. Accessed: Sept. 22, 2025. [Online]. Available: http://103.8.12.212:33180/unj/index.php/isc-beam/article/view/42678

[2] S. Susilawati, R. Falefi, and A. Purwoko, “Impact of COVID-19’s Pandemic on the Economy of Indonesia,” *Bp. Int. Res. Crit. Inst.-J. BIRCI-J.*, vol. 3, no. 2, pp. 1147–1156, 2020.

[3] Z. Abidin and T. Tobibatussa’adah, “The impact of covid-19 pandemic on education and judicial practice in indonesia,” *Riayah J. Sos. Dan Keagamaan*, vol. 5, no. 02, pp. 122–130, 2020.

[4] L. Octavia and R. Rachmalina, “Child malnutrition during the COVID-19 pandemic in Indonesia,” *Pediatr. Gastroenterol. Hepatol. Nutr.*, vol. 25, no. 4, p. 347, 2022.

[5] A. Y. Perdana, R. Latuconsina, and A. Dinimaharawati, “Prediksi Stunting Pada Balita Dengan Algoritma Random Forest,” *EProceedings Eng.*, vol. 8, no. 5, 2021.

[6] I. W. Pranata *et al.*, “Prevention of stunting through improving maternal parenting and early detection of pregnancy risk factors,” *J. Pengabdi. Masy. Bestari*, vol. 1, no. 9, pp. 1025–1034, 2022.

[7] L. H. Kusumawardani, U. Rachmawati, M. Jauhar, and I. G. A. P. D. Rohana, “Community-based stunting intervention strategies: Literature review,” *Dunia Keperawatan J. Keperawatan Dan Kesehat.*, vol. 8, no. 2, pp. 259–268, 2020.

[8] R. Setiyabudi, “Stunting, risk factor, effect and prevention,” *MEDISAINS J. Ilm. Ilmu-Ilmu Kesehat.*, vol. 17, no. 2, pp. 24–25, 2019.

[9] S. Widari, N. Bachtiar, and E. Primayesa, “Faktor Penentu Stunting: Analisis Komparasi Masa Millenium Development Goals (MDGs) dan Sustainable Development Goals (SDGs) di Indonesia,” *J. Ilm. Univ. Batanghari Jambi*, vol. 21, no. 3, pp. 1338–1346, 2021.

[10] J. R. Khan, J. H. Tomal, and E. Raheem, “Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh,” *Inform. Health Soc. Care*, vol. 46, no. 4, pp. 425–442, Dec. 2021, doi: 10.1080/17538157.2021.1904938.

[11] M. S. Haris, A. N. Khudori, and W. T. Kusuma, “Perbandingan metode supervised machine learning untuk prediksi prevalensi stunting di Provinsi Jawa Timur,” *J. Teknol. Inf. Dan Ilmu Komput. JTIIK*, vol. 9, no. 7, pp. 1571–1576, 2022.

[12] M. S. Haris, M. Anshori, and A. N. Khudori, “Prediction of stunting prevalence in east java province with random forest algorithm,” *J. Tek. Inform. Jutif*, vol. 4, no. 1, pp. 11–13, 2023.

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

[13] W. A. Amelia, “Implementasi Jaringan Syaraf Tiruan Sebagai Sistem Diagnosis Stunting Balita Menggunakan Metode Backpropagation Di Puskesmas Brebes,” *J. Minfo Polgan*, vol. 12, no. 2, pp. 1874–1883, 2023.

[14] I. K. A. Wiraguna, E. Setyati, and E. Pramana, “Prediksi Anak Stunting Berdasarkan Kondisi Orang Tua Dengan Metode Support Vector Machine Dengan Study Kasus Di Kabupaten Tabanan-Bali,” *SMATIKA J. STIKI Inform. J.*, vol. 12, no. 01, pp. 47–54, 2022.

[15] M. Amirudin and A. D. Wowor, “Analisis Perbandingan Klasifikasi Balita Beresiko Stunting Menggunakan Metode Support Vector Machine dan Decission Tree,” in *Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media*, 2023, pp. 581–591. Accessed: Sept. 22, 2025. [Online]. Available: https://conferences.ittelkom-pwt.ac.id/index.php/centive/article/download/217/146

[16] A. Y. Labolo, S. Mooduto, A. Bode, and I. C. R. Drajana, “Penerapan Algoritma Spport Vector Machine dan K-Nearest Neighbor Menggunkan Feature Selection Backward Elimination Untuk Prediksi Status Penderita Stunting Pada Balita,” *J. Tecnoscienza*, vol. 6, no. 2, pp. 374–388, 2022.

[17] H. G. Pangestu, R. Y. Sinaga, F. Z. Ulya, U. Athiyah, A. W. Muhammad, and F. Alameka, “Analisis Efisiensi Metode K-Nearest Neighbor dan Forward Chaining Untuk Prediksi Stunting Pada Balita,” *Inf. Mulawarman J Ilm Ilmu Komput*, vol. 18, no. 2, p. 78, 2023.

[18] M. Yunus, M. K. Biddinika, and A. Fadlil, “Classification of Stunting in Children Using the C4. 5 Algorithm,” *J. Online Inform.*, vol. 8, no. 1, pp. 99–106, 2023.

[19] S. Melyani, S. Z. Harahap, and I. Irmayanti, “Prediction of Stunting in Toddlers Combining the Naive Bayes Method and the C4. 5 Algorithm,” *Sink. J. Dan Penelit. Tek. Inform.*, vol. 8, no. 2, pp. 1160–1168, 2024.

[20] R. B. Abiyyi, E. R. Subhiyakto, and F. T. Sabilillah, “Centing: Aplikasi Cegah Stunting Anak berbasis Android menggunakan TensorFlow Lite,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 2, pp. 625–634, 2024.

[21] X. Gong and Z.-Z. Rao, “Construction of a model for analyzing the trend of suicidal tendency among college students based on mental health monitoring,” *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, 2024, doi: 10.2478/amns-2024-3479.

[22] M. H. Tito *et al.*, “Advancing vector-borne disease prediction through functional classifier integration: A novel approach for enhanced modeling,” *Lett. Anim. Biol.*, 2024, doi: 10.62310/liab.v4i1.135.

[23] C. Pabitha, V. Kalpana, E. S. Sv, A. Pushpalatha, G. Mahendran, and S. Sivarajan, “Development and Implementation of an Intelligent Health Monitoring System using IoT and Advanced Machine Learning Techniques,” *J. Mach. Comput.*, 2023, doi: 10.53759/7669/jmc202303037.

[24] T. N. Abiodun, D. Okunbor, and V. C. Osamor, “Remote Health Monitoring in Clinical Trial using Machine Learning Techniques: A Conceptual Framework,” *Health Technol.*, vol. 12, no. 2, 2022, doi: 10.1007/s12553-022-00652-z.

[25] T. Sugihartono and R. R. C. Putra, “Penerapan Metode Support Vector Machine Dalam Classifikasi Ulasan Pengguna Aplikasi Mobile JKN,” *SKANIKA Sist. Komput. Dan Tek. Inform.*, vol. 7, no. 2, 2024, doi: 10.36080/skanika.v7i2.3193.

[26] D. Damayanti, D. I. Efendi, D. Solihudin, C. L. Rohmat, and S. E. Permana, “Pemetaan Opini Publik Terhadap Perubahan Kebijakan BPJS Kesehatan Dengan Pendekatan Support Vector Machine(svm) Dalam Analisis Sentimen,” *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 1, 2024, doi: 10.36040/jati.v8i1.8304.

[27] N. Maulida, N. Suarna, and W. Prihartono, “Analisis Ulasan Sentimen Aplikasi Mobile JKN Dengan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization,” *JATI J. Mhs. Tek. Inform.*, vol. 8, no. 2, 2024, doi: 10.36040/jati.v8i2.9105.

[28] R. A. Raharjo, I. M. G. Sunarya, and D. G. H. Divayana, “Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin Covid-19 Di Twitter,” *Elkom J. Elektron. Dan Komput.*, vol. 15, no. 2, 2022, doi: 10.51903/elkom.v15i2.918.

Vol. 6, No. 4, Agustus 2025, Page. 2917-2928
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.4.4843

[29] G. Sanhaji, A. Febrianti, and H. Hidayat, "Aplikasi DIATECT Untuk Prediksi Penyakit Diabetes Menggunakan SVM Berbasis Web," *J. Tekno Kompak*, vol. 18, no. 1, 2024, doi: 10.33365/jtk.v18i1.3643.

[30] A. Jalil, A. Homaidi, and Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2070–2079, 2024.

[31] A. F. Najwa and P. H. Gunawan, "Enhancing Stunting Prediction for Indonesian Children Using Machine Learning with SMOTE Data Balancing," *J. Comput. Sci. Eng.*, vol. 18, no. 4, pp. 203–213, 2024.

[32] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bull. Inf. Technol. BIT*, vol. 5, no. 2, pp. 56–64, 2024.

[33] T. Sugihartono, B. Wijaya, M. Marini, A. P. Alkayess, and H. A. Anugerah, "Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 667–682, 2025.

[34] O. Pahlevi, D. A. N. Wulandari, L. K. Rahayu, H. Leidiyana, and Y. Handrianto, "Model Klasifikasi Risiko Stunting Pada Balita Menggunakan Algoritma CatBoost Classifier," *Bull. Comput. Sci. Res.*, vol. 4, no. 6, pp. 414–421, 2024.