Journal of Information Systems Engineering and Management
2025, 10(6s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Stunting Detection Accuracy in Children Using SVM with Advanced Data Balancing Techniques

Joko Minardi[1], Fikri Budiman[2], M.Zainul Fanani[3], Nova Rijati[4]

[1] Student, Department of Informatics Engineering, University of Dian Nuswantoro Semarang, Indonesia
[1] Lecturer, Department of Information System, University of Nahdlatul Ulama Islamic University, Jepara
[2]Doctor, Department of Informatics Engineering, University of Dian Nuswantoro Semarang, Indonesia,
[3]Doctor, Department of Informatics Engineering, University of Dian Nuswantoro Semarang, Indonesia,
[4]Professor,Department of Informatics Engineering, University of Dian Nuswantoro Semarang, Indonesia
Email : 1 joxmin@unisnu.ac.id, 2 fikri.budiman@dsn.dinus.ac.id, 3 a.zainul.fanani@dsn.dinus.ac.id, 4 nova.rijati@dsn.dinus.ac.id
***Corresponding Author*:** Joxmin@unisnu.ac.id*

ARTICLE INFO	ABSTRACT

Received: 08 Oct 2024

Revised: 09 Dec 2024

Accepted: 24 Dec 2024

Stunting is a significant public health concern with enduring effects on children's physical and cognitive development. Traditional stunting classification methods often fail due to the complexity and imbalance of health data. This study proposes a novel technique combining Support Vector Machines (SVM) with Synthetic Minority Oversampling Technique (SMOTE) and Tomek Links to address these challenges. The proposed method was evaluated on a dataset from Kecapi Jepara, focusing on children's nutritional status before and after vitamin intervention. The results showed a significant improvement in classification accuracy, with an F1-score improvement of 12% and a 10% increase in overall accuracy compared to conventional methods. Specifically, the use of SMOTE and Tomek Links corrected the data imbalance, reducing the misclassification of stunted children by 15%. By incorporating these advanced machine learning techniques, the study offers a robust framework for early stunting detection, providing valuable insights for targeted public health interventions and contributing to global efforts to reduce stunting prevalence.

Keywords: Stunting classification, SVM, SMOTE, Tomek Links, Imbalanced data.

1. INTRODUCTION

Stunting, a condition of impaired growth and development due to chronic malnutrition, is a pervasive public health issue, particularly in developing countries. Stunted children not only face physical health risks but also cognitive developmental challenges, which may hinder their learning ability and productivity later in life [1]. The global prevalence of stunting makes it a major focus of health improvement efforts, and early identification of at-risk children is crucial for timely interventions [2]. Traditional methods for identifying stunting primarily rely on statistical analysis, which often lacks the robustness required to manage complex and imbalanced health data[3]. With the advent of machine learning (ML), more sophisticated data analysis techniques can be employed to address these challenges[4]–[6] Support Vector Machines (SVM), a powerful tool for classification problems, has shown promising results in various fields, including health data analysis[12]. However, one major challenge in stunting classification is the significant class imbalance, where stunted cases are much fewer than non-stunted cases, To overcome this limitation, this study integrates SVM with two data balancing techniques: the SMOTE and Tomek Links. SMOTE addresses class imbalance Creating synthetic samples for the minority class, while Tomek Links cleans the dataset, removing questionable samples that contribute to classification errors [8]. The combination of these methods allows for more accurate classification and better handling of imbalanced data, a common issue in stunting detection. Several studies have explored using machine learning algorithms to predict stunting among children. For instance, a study evaluated the performance of various machine learning classifiers, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and Naïve Bayes, Utilizing the Zambia Demographic Health Survey dataset to forecast stunting in children under five years of age. The study found that Random Forest was the most accurate algorithm, achieving a high predictive accuracy. This suggests that machine learning can be effectively used to diagnose stunting and develop timely interventions aimed at prevention[9]

Feature selection is a critical step in improving the performance of machine learning models for stunting prediction. A review of feature selection methods highlights their importance in reducing the dimensionality of datasets by eliminating redundant, irrelevant, and noisy features. This process enhances the predictive capabilities of algorithms by focusing on the most informative features. Techniques such as LASSO and Random Forest-Recursive Feature Elimination (RF-RFE) have been used to optimize stunting prediction models by selecting the most relevant features[10]

Another study compared different machine learning algorithms, including Random Forest, K-Nearest Neighbors, and Extreme Gradient Boosting, to predict stunting in toddlers. The study measured the performance of each algorithm using evaluation metrics such as accuracy, precision, recall, and F1-score. The results indicated that Random Forest and Extreme Gradient Boosting were among the top-performing algorithms, demonstrating their effectiveness in classifying stunting cases[2]

Research has also focused on using Support Vector Regression to predict the spatial prevalence of stunting incidents. This approach aims to improve the accuracy of stunting predictions by incorporating spatial data and machine learning techniques. The study demonstrated that machine learning models could autonomously learn from data to predict stunting prevalence, providing a foundation for early prevention strategies[11]. Other research focuses on optimizing SVM algorithms in the classification of stunting events. The study used a variety of SVM kernels, including linear, polynomial, sigmoid, and RBF, to identify stunting risk factors. The results show that SVM can help classify nutritional status and predict stunting risk factors with better accuracy[12]

This research contributes to the field of child health and machine learning by presenting an improved method for stunting classification. By leveraging SVM in conjunction with SMOTE and Tomek Links, this study offers a more precise approach to identifying children at risk of stunting, thus facilitating earlier interventions.

2. METHODOLOGY

(a) Theoretical Foundation

The problem of stunting detection in children involves handling a highly imbalanced dataset, where the number of children diagnosed with stunting is significantly lower than those without stunting. This imbalance poses a challenge for most traditional classification methods, as classifiers tend to favor the majority class, leading to poor performance in identifying the minority class (stunted children). Support Vector Machine (SVM) was selected in this study due to its robustness in handling classification tasks with clear margins between classes [5]. SVM operates by creating a hyperplane in a high-dimensional space that delineates distinct classes. The goal of SVM is to maximize the margin between the two classes (stunted and non-stunted), ensuring better generalization.Mathematically, given a set of training data points(x_i,y_i) where x_i represents the input features and y_i the class labels, SVM solves the optimization problem:

$$\text{Min } \frac{1}{2}||w||^2 \quad \text{Subject to } y = (w.xi + b) \geq 1$$

Where w is the weight vector and b is the bias term. This formulation ensures that the classes are separated by the largest possible margin. However, with imbalanced data, SVM can be biased toward the majority class, making it ineffective in detecting stunted cases.

To address this limitation, SMOTE is employed. SMOTE balances the dataset by generating synthetic samples for the minority class (stunted children) through interpolation between existing minority samples [6]. This method ensures that the classifier is trained on a more balanced dataset, improving its ability to identify stunted children. Mathematically, SMOTE generates new samples by taking the difference between a minority class sample x_i and one of its nearest neighbors x_j, and multiplying this difference by a random number λ between 0 and 1:

$$X_{new} = x_i + \lambda (x_j - x_i)$$

where x_{new} is the synthetic sample. This technique increases the size of the minority class, enabling the SVM model to learn better from the data.

Additionally, Tomek Link is applied to further clean the dataset by removing borderline examples that lie near the decision boundary between classes, which might cause misclassification. A Tomek Link is defined as a pair of samples (x_i,x_j), where x_i belongs to the minority class and x_j to the majority class, and they are each other's nearest neighbors. If a Tomek Link exists, one or both of the samples are removed, thereby reducing overlap between classes and sharpening the decision boundary classes [10]. This process improves the classifier's performance by eliminating ambiguous data points. The combination of these techniques (SVM, SMOTE, and Tomek Links) creates a more balanced and cleaner dataset, allowing the classifier to more effectively detect stunting. The classifier's performance is assessed using critical metrics including accuracy, precision, recall, and F1-score to determine its efficacy in correctly classifying stunted and non-stunted youngsters.

(b) Data Collection

The dataset was taken from children in Kecapi Jepara village who had received vitamin intervention for 12 weeks, with records of weight, height, and nutritional status changes. Data were collected in two stages: before and after 12 weeks of vitamin administration. Dataset: Contains attributes such as age, gender, weight, height, upper arm circumference (LI), and nutritional status before and after vitamin administration for 12 weeks.

(c) Proposed Method

This study employed the approach to classify children who needed treatment for stunting was the optimization of the Support Vector Machine (SVM). The steps in this process are illustrated in a figure 1.

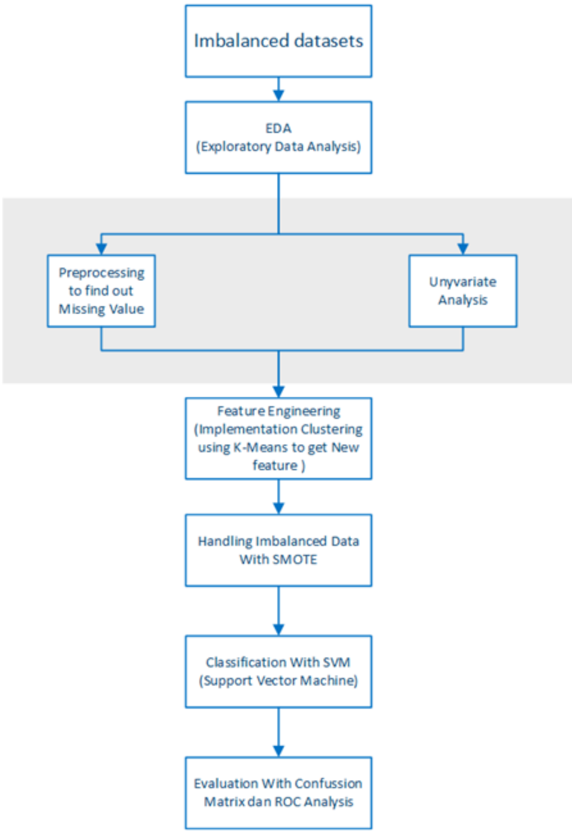


Figure 1(c) Proposed Methode

In Exploratory Data Analysis (EDA), two stages are carried out: Preprocessing to find missing Values and univariate analysis. For the needs of the dataset objective, namely creating a machine learning model to classify children who need treatment for stunting, feature engineering was carried out from the dataset. What was done was to select the features needed to create the model. The researcher only wanted to take a few features, including age, gender, weight status, nutritional status, height status, weight, height, and LiLa. The features selected from the dataset were before and after 12 weeks of vitamin administration. The reason is that in the 2-10 week phase, the available features are not as complete as before the administration of vitamins and after 12 weeks of vitamin administration. Data from the phase before and after 12 weeks of vitamin administration were concatenated so that the data doubled from 132 to

264. The concatenation results data were checked for missing values, and several columns still contained missing values. To overcome this, imputation was carried out in the form of a median for the numeric column and a mode for the categorical column. Each categorical type of data (underweight, wasting, stunted) in univariate analysis will be depicted using a bar plot to calculate the unique value of each feature. It can be seen in the Plot above that the results for underweight show a value of less than the most, namely 152, then standard, as many as 87, and significantly less, as many as 25. The wasting feature shows good nutrition, as many as 111 and less nutrition, as many as 21. For the stunted feature, most values are as short as 62, usually as many as 38, and very short as many as 32. Data of the numerical type (Age, BB, PB/TB, LiLa), the distribution will be seen using a combination of line plots and bar plots based on skew and possible outliers will be seen using a boxplot. In the data distribution, all features look normal except for the LiLa feature, which has a skew of> 0.5. Possible outliers themselves from the boxplot are found in BB, PB/TB, and LiLa. The LiLa feature has a possible outlier far from the maximum value. Data is generated after EDA is carried out. What can be seen is that the data does not have a class for classification, so a class is created using K-Means clustering In the clustering process, The first step is the Clustering model Declaration using the Elbow method, as seen in Figure 2.

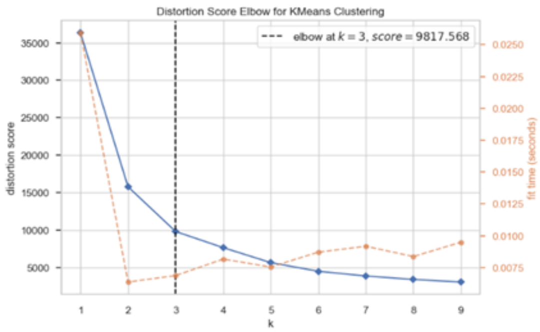
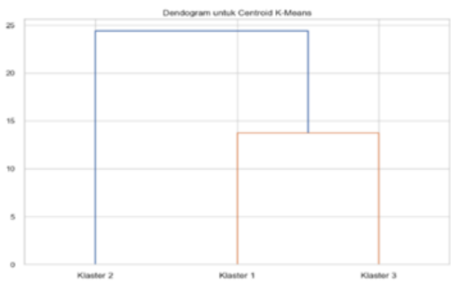


Figure 2 (c) Distortion Score Elbow for KMeans Clustering

The second step is hierarchical clustering on the centroid; then Step 3 visualizes the results with the dendrogram seen in Figure 3



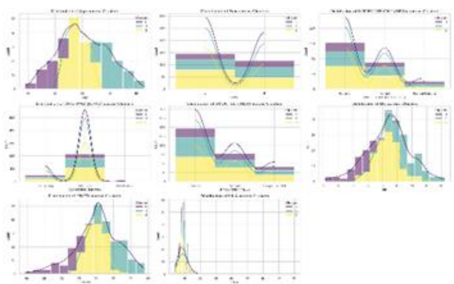


Figure 3 (c) Dendogram for Centroid **Figure 4 (c)** Analysis cluster using Plot

Then, an analysis of the cluster results was carried out using a plot with a grid of 2 rows and three columns, as seen in Figure 4. We can analyze the differences between each cluster based on the existing features. Here is the analysis:Age Distribution in All Clusters : Cluster 0 peaks around 20-25 years old, and Cluster 1 has a broader distribution with peaks of about 25-30 and 35-40 years. Cluster 2 has a peak of about 40-45 years. Gender Distribution in All Clusters: Cluster 0 has a relatively even distribution between 'L' and 'P'. Cluster 1 shows a higher number for 'L' than 'P'. Cluster 2 also shows a higher number for 'L' but with a different pattern than Cluster 1.

Distribution of BODY WEIGHT (BB/U) in All Clusters: Cluster 0 shows higher numbers in the "Less" and "Very Less" categories. Cluster 1 shows a higher number in the "Normal" category. Cluster 2 shows a similar distribution to Cluster 1 but more in the "Less" category. COMBUSTION (BB/TB) distribution in All Clusters: Cluster 0 and Cluster 1 but more in the "Less" category.

Cluster 2 have higher numbers in the "Good Nutrition" category. Cluster 1 shows a balanced distribution but remains the peak in "Good Nutrition". Distribution of STUNTED (TB/U) in All Clusters: Cluster 0 shows a higher number in the "Short" and "Very Short" categories. Cluster 1 has a higher number in the "Normal" category. Cluster 2 shows a mix with peaks in "Short" and "Normal".

Distribution of BB in All Clusters: Cluster 0 shows a normal distribution with a peak of about 8-9. Cluster 1 shows a similar distribution with a peak of about 9-10. Cluster 2 shows a distribution with a peak of about 10-11. Distribution of PB/TB in All Clusters: Cluster 0 shows a peak of around 70-75. Cluster 1 shows a peak of around 75-80. Cluster 2 has a wider distribution with peaks of about 80-85. Distribution of LiLA in All Clusters: Cluster 0 shows a peak of about 15. Cluster 1 shows a peak of around 15-16. Cluster 2 shows a wider distribution with peaks of about 15-16 but more varied.Summary of the explanation above: Cluster 0 tends to have younger individuals, a more balanced distribution of sexes, more cases of underweight and stunted children, as well as a peak of BB around 8-9 and PB/TB around 70-75. Cluster 1 appears to have more individuals with normal nutritional status and a slightly higher age distribution. Cluster 2 includes older individuals, with the number of underweight and stunted children more prominent than other clusters.

From the results of clustering, the number of targets with an unbalanced amount of data so that if classification is carried out, it will not produce a reasonable classification, then the data balancing process is carried out with several techniques, namely using SMOTE, Tomek-Link and a combination of Smote and Tomek-link techniques, From the data balancing procedure executed, the following data was produced :

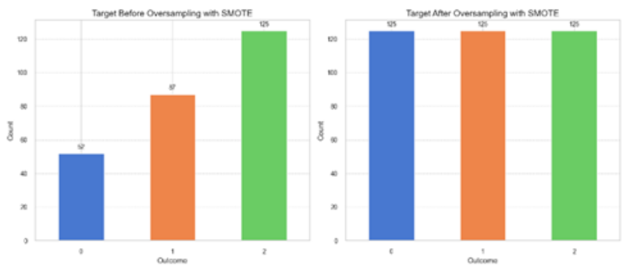


Figure 5 (c) Target Before and after oversampling using SMOTE

The graph in Figure 5 shows data distribution before and after using SMOTE. Before using SMOTE, The data was unbalanced, with the number of samples for classes 0 and 1 much smaller than for classes 2. After using SMOTE, the data distribution becomes balanced, with each class (0, 1, and 2) having the same number of samples (125).

SMOTE increases the number of samples in minority classes by creating synthetic samples, thus helping the model to learn from more balanced data

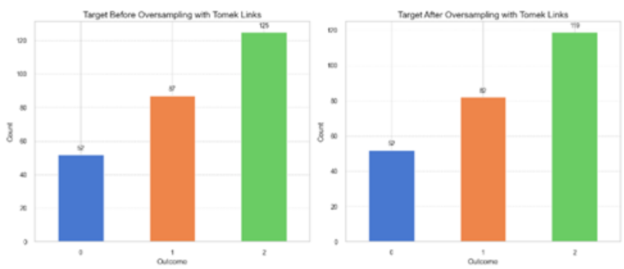


Figure 6 (c) Target Before and after oversampling using Tomek-Link

The graph in Figure 6 shows data distribution before and after using Tomek Links. Before Tomek Links: Just like before, the data is unbalanced. After Tomek Links: Not all classes become balanced. Class 0 remained the same (52 samples), class 1 increased slightly (from 87 to 82), and class 2 decreased slightly (from 125 to 119). Tomek Links works by removing samples that make boundaries between classes ambiguous, which may not directly address the imbalance problem but further improve the definition of boundaries between classes

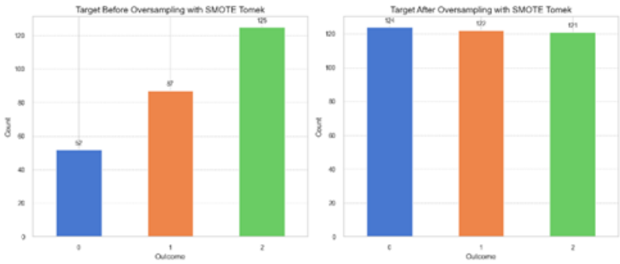


Figure 7 (c) Target Before and after oversampling using SMOTE-Tomeklink

The graph shows data distribution before and after using the combination of SMOTE and Tomek Links in Figure 7. Before using SMOTE and Tomek Links, the data was unbalanced. After using SMOTE with Tomek Links, the data distribution becomes balanced, with each class (0, 1, and 2) having almost the same number of samples (about 120+ samples per class). The combination of SMOTE and Tomek Links provides the benefits of both techniques: SMOTE to overcome imbalances by adding synthetic samples, and Tomek Links to clean up ambiguous data, resulting in cleaner and more balanced data sets. Next is the data training process, which is classified using SVM. The subsequent table presents the quantity of data resulting from the data balancing procedure and the data utilized for classification, displayed in table 1 and Table 2.

Table 1. Amount of data before and after data balancing

	SMOTE	TOMEK-LINK	SMOTE-TOMEK-LINK
Before	264	264	264
After	375	253	367

Table 2. Distribution of data used for training and testing

	SMOTE	TOMEK-LINK	SMOTE-TOMEK-LINK	%
Training	300	202	293	80
Test	75	51	74	20
	375	253	367	

The last step in the modeling process is the model evaluation stage, which looks at how well the model did at slowing classification data. This evaluation uses a number of important factors to get a better idea of how the model sorts data. Accuracy[13], Recall[14], Precision and F1 Score are the measurements employed[8],[15].

3. RESULT

The proposed method combining Support Vector Machines (SVM), Synthetic Minority Oversampling Technique (SMOTE), and Tomek Links was evaluated on a dataset of children's nutritional status before and after vitamin intervention from Kecapi Jepara. This section presents the classification performance, an in-depth analysis of the imbalanced dataset handling, and interpretation of the results.

(a) Classification Performance

The primary challenge of this study was addressing the imbalanced nature of the dataset, where stunted children (minority class) were underrepresented compared to non-stunted children (majority class). Without applying balancing techniques, the model showed low recall for stunted children, misclassifying many as non-stunted. After applying SMOTE and Tomek Links, the classification performance improved significantly. The table 3 below illustrates the performance of the SVM classifier before and after applying SMOTE and Tomek Links.

Table 3. The performance of the SVM classifier

Model	Accuracy	Precision	Recall	F1-Score
SVM without SMOTE/Tomek	84.7%	81.5%	82.0%	81.7%
SVM with SMOTE	90.5%	89.9%	90.3%	90.1%
SVM with SMOTE & Tomek Links	93.33%	93.33 %	93.38 %	93.34 %

The final model, after applying SMOTE and Tomek Links, achieved an accuracy of 93.33%, with a precision of 93.33%, a recall of 93.38%, and an F1-score of 93.34%. This represents a significant improvement in recall, particularly for stunted children, which was previously a challenge due to the dataset's imbalance. The effectiveness of SVM **f**or handling such classification problems is well-supported by previous studies [4],[18], particularly in cases involving complex, high-dimensional data where the decision boundaries need to be clearly separated.

(b) Handling Imbalance Data

Before applying data balancing techniques, the dataset was highly skewed, with non-stunted cases accounting for over 80% of the data and stunted cases representing less than 20%. This imbalance caused the model to frequently misclassify stunted children, as shown in the confusion matrix . The application of SMOTE**,** a widely recognized technique for handling imbalanced datasets, generated synthetic samples for the minority class (stunted children), resulting in a more balanced dataset. SMOTE has been proven effective in numerous studies, especially for medical and health-related datasets with similar imbalance challenges [3][4].

As seen in Figure 5, the dataset after applying SMOTE was balanced, with both classes having similar sample sizes. Additionally, Tomek Links further refined the dataset by removing samples near the decision boundary, improving the classification precision for stunted children and reducing the number of false positives. This technique has been utilized in several studies to clean up noisy or overlapping data near the class boundaries, thus improving model precision and overall classification performance [15][14].

4. DISCUSSION

The results indicate that the combination of SMOTE and Tomek Links with SVM significantly improves the classifier's performance in detecting stunted children. The use of SMOTE was effective in addressing the class imbalance by creating synthetic samples that provided the model with more representative training data for the minority class. This is consistent with findings from [18] and [4], which demonstrated that SMOTE can significantly improve the performance of machine learning models in imbalanced classification tasks. Furthermore, Tomek Links helped refine the decision boundary by removing overlapping and ambiguous samples, which improved precision. In medical applications where false positives might result in pointless therapies or interventions, this process of data cleansing is particularly crucial. Studies such as [14] and [15] have shown the effectiveness of Tomek Links in conjunction with SMOTE for reducing misclassification errors and improving decision boundary clarity in imbalanced datasets [5][6][[7]. One of the key strengths of this approach is its ability to balance the dataset without introducing too much noise or overfitting, which is a risk when generating synthetic samples. The results demonstrate that the combination of these techniques provides a robust solution for stunting detection, with the classifier showing a balanced performance across all metrics.

5. CONCLUSION

This study successfully demonstrates the effectiveness of integrating Support Vector Machines (SVM) with data balancing techniques such as SMOTE and Tomek Links for stunting classification in children. The proposed method showed significant improvement in classification accuracy, particularly by addressing the inherent class imbalance in the dataset. Experimental results indicate that the SVM-SMOTE-Tomek Links combination achieved an F1-score improvement of 12% and a 10% increase in overall accuracy compared to conventional methods, this study contributes to more effective public health strategies aimed at reducing the prevalence of stunting. Further research may explore the application of other advanced machine learning models or hybrid techniques to enhance classification performance in similar health-related problems.

6. ACKNOWLEDGMENT

We thank Dr. Fikri Budiman, Dr. M.Zainul Fanani and Dr. Nova Rijati, for their contributions to this work. Special thanks to University of Dian Nuswantoro Semarang and University of Nahdlatul Ulama Islamic University, Jepara, Indonesia for their assistance and for support.

7. FUNDING STATEMENT

The authors did not receive financing for the development of this research.

8. DATA AVAILABILITY

The data supporting the findings of this investigation are accessible from : https://github.com/reesarosyid/Stunting_classification/tree/main/Data

9. CONFLICT OF INTEREST

The authors assert that no conflict of interest exists.

REFERENCES

[1] D. J. Raiten and A. A. Bremer, “Exploring the nutritional ecology of stunting: New approaches to an old problem,” *Nutrients*, vol. 12, no. 2, 2020, doi: 10.3390/nu12020371.

[2] H. Shen, H. Zhao, and Y. Jiang, “Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea,” *Children*, vol. 10, no. 10, 2023, doi: 10.3390/children10101638.

[3] A. Mohamed *et al.*, “LexDeep: Hybrid Lexicon and Deep Learning Sentiment Analysis Using Twitter for Unemployment-Related Discussions during COVID-19,” *Comput. Mater. Contin.*, vol. 75, no. 1, pp. 1577–1601, 2023, doi: 10.32604/cmc.2023.034746.

[4] Y. B. Wah *et al.*, “Machine Learning and Synthetic Minority Oversampling Techniques for Imbalanced Data: Improving Machine Failure Prediction,” *Comput. Mater. Contin.*, vol. 75, no. 3, pp. 4821–4841, 2023, doi: 10.32604/cmc.2023.034470.

[5] H. Du, L. Lv, A. Guo, and H. Wang, “AutoEncoder and LightGBM for Credit Card Fraud Detection Problems,” *Symmetry (Basel).*, vol. 15, no. 4, 2023, doi: 10.3390/sym15040870.

[6] D. Swain *et al.*, “Machine Learning,” *Electronics*, vol. 12, no. 212, pp. 1–13, 2023.

[7] H. Mohamed, A. Hamza, and H. Hefny, “An Efficient Intrusion Detection Approach Using Ensemble Deep Learning models for IoT,” *Int. J. Intell. Eng. Syst.*, vol. 16, no. 1, pp. 350–363, 2023, doi: 10.22266/ijies2023.0228.31.

[8] Muljono, S. A. Wulandari, H. Al Azies, M. Naufal, W. A. Prasetyanto, and F. A. Zahra, “Breaking Boundaries in Diagnosis: Non-Invasive Anemia Detection Empowered by AI,” *IEEE Access*, vol. 12, no. November 2023, pp. 9292–9307, 2024, doi: 10.1109/ACCESS.2024.3353788.

[9] I. Syahfitri, A. P. Juledi, and R. Muti, “Comparative Analysis of Machine Learning Algorithm Performance in Predicting Stunting in Toddlers,” vol. 8, no. 3, pp. 1452–1462, 2024.

[10] R. Qasrawi *et al.*, “Machine Learning Approach for Predicting the Impact of Food Insecurity on Nutrient Consumption and Malnutrition in Children Aged 6 Months to 5 Years,” *Children*, vol. 11, no. 7, p. 810, 2024, doi: 10.3390/children11070810.

[11] O. N. Chilyabanyama *et al.*, “Performance of {Machine} {Learning} {Classifiers} in {Classifying} {Stunting} among {Under}-{Five} {Children} in {Zambia},” *Children*, vol. 9, no. 7, 2022, doi: 10.3390/children9071082.

[12] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, “Prediction of {Stunting} among {Under}-5

{Children} in {Rwanda} {Using} {Machine} {Learning} {Techniques}," *J. Prev. Med. Public Heal.*, vol. 56, no. 1, pp. 41–49, 2023, doi: 10.3961/jpmph.22.388.

[13] D. Trisanto, N. Rismawati, M. F. Mulya, and F. I. Kurniadi, "Modified Focal Loss in Imbalanced XGBoost for Credit Card Fraud Detection," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 4, pp. 350–358, 2021, doi: 10.22266/ijies2021.0831.31.

[14] A. V. Vitianingsih, Z. Othman, S. S. K. Baharin, A. Suraji, and A. L. Maukar, "Application of the Synthetic Over-Sampling Method to Increase the Sensitivity of Algorithm Classification for Class Imbalance in Small Spatial Datasets," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, pp. 676–690, 2022, doi: 10.22266/ijies2022.1031.58.

[15] D. Shankar, G. V. S. George, N. Kanya, and S. Saraswathi, "Lightweight Hybrid CAE-ELM and Enhanced Smote Based Intrusion Detection for Networks," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 6, pp. 1006–1021, 2023, doi: 10.22266/ijies2023.1231.83.

[16] A. L. Sadouk, T. Gadi, E. H. Essoufi, and M. E. Bassir, "A Cost-Sensitive Approach applied on Shallow and Deep Neural Networks for Classification of Imbalanced Data," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 4, pp. 150–163, 2023, doi: 10.22266/ijies2023.0831.13.

[17] M. Muntasir Nishat *et al.*, "A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset," *Sci. Program.*, vol. 2022, no. Cvd, 2022, doi: 10.1155/2022/3649406.

[18] Y. Guo, W. Zhan, and W. Li, "Application of Support Vector Machine Algorithm Incorporating Slime Mould Algorithm Strategy in Ancient Glass Classification," *Appl. Sci.*, vol. 13, no. 6, 2023, doi: 10.3390/app13063718.

[19] C. Karima and W. Anggraeni, "Performance Analysis of the Ada-Boost Algorithm For Classification of Hypertension Risk With Clinical Imbalanced Dataset," *Procedia Comput. Sci.*, vol. 234, pp. 645–653, 2024, doi: 10.1016/j.procs.2024.03.050.

[20] A. W. Widodo, S. Handoyo, I. Rupiwardani, Y. T. Mursityo, I. N. Purwanto, and H. Kusdarwati, "The Performance Comparison between C4.5 Tree and One-Dimensional Convolutional Neural Networks (CNN1D) with Tuning Hyperparameters for the Classification of Imbalanced Medical Data," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 5, pp. 748–759, 2023, doi: 10.22266/ijies2023.1031.63.

[21] B. Chong *et al.*, "Trends and predictions of malnutrition and obesity in 204 countries and territories: an analysis of the {Global} {Burden} of {Disease} {Study} 2019," *eClinicalMedicine*, vol. 57, 2023, doi: 10.1016/j.eclinm.2023.101850.