JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 10, No. 3, July 2026, pp. 768-782

Application of Ensemble Bagging Support Vector Machine for Early Detection of Childhood Stunting

Alfiyah Hanun Nasywa[1], Solimun[1*], Achmad Efendi[1], Adji Achmad Rinaldo Fernandes[1], Celia Sianipar[1], Fachira Haneinanda Junianto[2]
[1]Department of Statistics, Universitas Brawijaya, Indonesia
[2]Department of Mathematics, Universitas Brawijaya, Indonesia
solimun@ub.ac.id

Article History:
Received : 27-10-2025
Revised : 05-12-2025
Accepted : 07-12-2025
Online : 01-07-2026

Keywords:
Stunting;
Support Vector Machine;
Kernel Polinomial;
Ensemble Bagging;
Classification.

ABSTRACT

Stunting is a significant public health issue in Indonesia, characterized by a child's height being below the age standard. Maternal knowledge and family economic level are key factors influencing children's nutritional status, thus requiring accurate classification methods for early stunting risk detection. This study aims to develop a machine learning-based classification model for stunting risk using Support Vector Machine (SVM) with a quadratic polynomial kernel and evaluate its performance improvement through the ensemble Bagging SVM approach. Primary data were collected from 100 mothers of children under five, using a five-point Likert scale questionnaire to assess maternal knowledge (X_1) and family economic level (X_2). The SVM model was constructed using a quadratic polynomial kernel and compared to Bagging SVM, which applies bootstrap resampling and majority voting. Model performance was evaluated using accuracy, sensitivity, and specificity. The basic SVM model yielded 85% accuracy, 90% sensitivity, and 80% specificity. The SVM Bagging approach showed performance improvements, with 95% accuracy, 100% sensitivity, and 94% specificity. These results indicate that SVM Bagging reduces misclassification. The SVM Bagging approach was more effective than a single SVM in classifying stunting risk. The novelty and scientific contribution of this study lie in applying ensemble machine learning methods, particularly Bagging SVM, to enhance early detection of stunting risk. This method offers a reliable solution for improving stunting risk classification accuracy and strengthening targeted nutrition interventions in Indonesia.

https://doi.org/10.31764/jtam.v10i3.36005

A. INTRODUCTION

Stunting is a significant public health issue due to its long-term effects on cognitive ability, educational attainment, economic productivity, and its role in perpetuating the cycle of poverty across generations (Putri et al., 2024). This condition not only hinders physical growth but also negatively impacts cognitive development, long-term productivity, and increases the risk of degenerative diseases in adulthood (WHO, 2020). According to recent data (Ministry, 2025), the national prevalence of stunting in Indonesia stands at 19.8%, with Malang Regency reporting a slightly lower rate of 16.4%. Despite various national programs, these figures remain above the target set by Indonesia's National Medium-Term Development Plan (RPJMN), which aims to reduce stunting by 14% by 2024 (Ministry, 2025). This indicates a significant challenge in addressing the root causes and effectively identifying stunting risks.

The urgency of addressing stunting is amplified by its far-reaching impacts, not only as a public health issue but also in terms of its effect on the quality of human capital in the future. Children affected by stunting are at greater risk for developmental delays and decreased productivity in adulthood (Victora et al., 2021). The causes of stunting are complex and influenced by geographical, socio-economic, and lifestyle factors that vary across regions. Identifying and modeling these diverse factors is critical for targeted intervention programs (Laksono et al., 2022). In Indonesia, stunting remains persistently high and is closely linked to socio-economic disparities, which affect households' access to adequate nutrition (Handayani et al., 2023).

Early detection of stunting risk is crucial for enabling timely interventions and ensuring that stunting prevention programs are effective (UNICEF, 2023). However, the main challenge in analyzing stunting data is the imbalance between the number of normal and stunted children, leading to biased results when using conventional classification algorithms (He & Garcia, 2009). Previous studies have attempted to address this imbalance, but they have often struggled to improve classification accuracy, especially for the minority class, which constitutes the majority of the problem. To overcome these limitations, this study proposes the use of an ensemble bagging Support Vector Machine (SVM) model to enhance the classification of stunting risk. SVM is known for its ability to create an optimal hyperplane that separates classes with maximum margins, and it performs well with high-dimensional data and non-linear relationships through the use of kernel functions, such as polynomial kernels. However, SVM’s performance can be compromised in the case of imbalanced data, as it tends to favor the majority class. This is where the ensemble learning approach becomes useful. Ensemble learning aggregates the predictions of multiple base learners, which improves the model’s stability and accuracy, particularly in imbalanced datasets (Breiman, 1996).

The novelty of this study lies in combining the ensemble bagging technique with SVM, addressing both the imbalance in stunting data and the non-linear nature of the relationship between predictors and stunting risk. Recent research by Karthikeyan & Ravichandran (2022) demonstrated that integrating Support Vector Machine (SVM) with multi-objective ensemble bagging optimization effectively enhances classification performance when dealing with imbalanced datasets. Similarly, Wattimena et al. (2023) developed an ensemble resampling-based SVM framework for multiclass imbalanced data, showing significant improvements in predictive precision. Furthermore, Rumagit & Pangkey (2021) revealed that bagging-based ensemble analysis contributes to greater model stability and reduces variance in classification tasks. Based on this background, this study focuses on developing the ensemble bagging SVM model for the classification of stunting risk in children, aiming to produce a more robust and accurate model to support early detection and intervention efforts for stunting in Indonesia.

B. METHODS

1. Research Design

This study aims to classify stunting risk using an ensemble bagging Support Vector Machine (SVM) model. The research follows a quantitative approach, focusing on machine learning techniques to address the challenges posed by imbalanced data in stunting risk classification. The main steps in the research are data collection, preprocessing, kernel selection, model development, ensemble bagging application, and performance evaluation.

2. Research Data

This study used primary data involving 100 mothers with children under five in Dadapan Village, Wajak District, Malang Regency. Samples were taken using quota sampling techniques and data were collected through questionnaires with a five-scale Likert scale. The variables analyzed included Maternal Knowledge (X_1), Economic Level (X_2), and Physical Stunting (Y). The Y variable as a binary-scale dependent variable (stunting and non-stunting) is used to build a classification model through the ensemble bagging approach. To maintain the validity of the data, the following inclusion and exclusion criteria are set.

a. Inclusion criteria
   1) My mother lives in Dadapan Village, Wajak District, Malang Regency.
   2) Have children aged 0–59 months under five.
   3) Be willing to participate in the research by signing an informed consent.
   4) Able to understand the instructions for filling out the questionnaire.

b. Exclusion criteria
   1) Mothers with children under five who have severe congenital diseases or congenital abnormalities that can affect growth status.
   2) Respondents who were not at the research site at the time of data collection.
   3) The questionnaire filled out is incomplete or invalid.

With these criteria, it is hoped that the selected respondents are truly representative of the population of mothers with toddlers in the research area, so that the results of the stunting risk classification can be obtained more accurately and can be scientifically accounted for.

3. Data preprocessing

The raw data was preprocessed to ensure it was clean and ready for model training. This involved handling missing values, standardizing numerical variables, and encoding categorical variables where necessary. Data splitting was performed, dividing the dataset into training (80%) and testing (20%) sets. Cross-validation was applied during model training to assess model stability and avoid overfitting.

4. Data analysis methods

In this study, we employed a combination of Support Vector Machine (SVM), Ensemble Bagging, and classification accuracy measures to predict stunting risk. The method is designed to address the challenge of imbalanced datasets and to enhance the predictive power of classification models.

a. Support Vector Machine (SVM)

SVM is a classification method that looks for an optimal linear separator hyperplane in that dimension, called a decision boundary, to separate the tuple from one class from another (Gao et al., 2025). This method can also be extended to non-linear cases through kernel functions so that it can provide better classification results than conventional methods (Talakua & Tomasouw, 2023). Data from the two classes can be separated by a hyperplane by using precise non-linear mapping to sufficiently high dimensions (Hattle et al., 2022). Support Vector Machine (SVM) applies the principle of structural risk minimization, which enables the model to achieve high accuracy and generalization capability, making it widely used for both classification and optimization problems (Dongoran & Cipta, 2025). Here is a description of the two classes separated by the hyperplane (Liu et al., 2023), as shown in Figure 1.

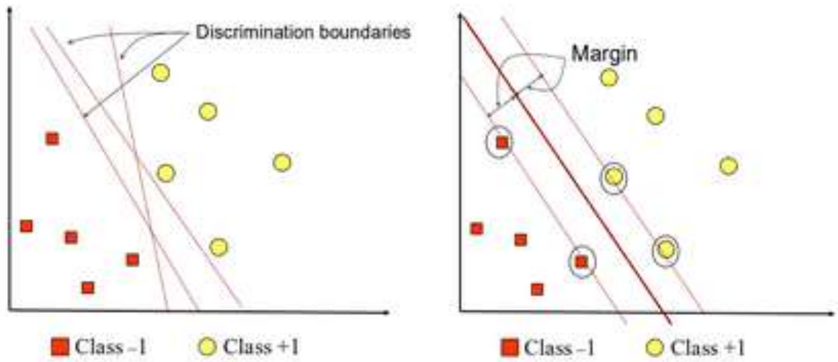

Figure 1. Hyperplane that separates the two classes

Suppose each data is denoted as $(\underset{\sim}{x}_i, \underset{\sim}{y}_i)$, where $\underset{\sim}{x}_i$ the vector variable is a predictor and $\underset{\sim}{y}_i \in \{-1, +1\}$ is a class label (Gao et al., 2025). The two classes are said to be perfectly separable if there is a hyperplane with an Equation (1).

$$\underset{\sim}{w}^T . \underset{\sim}{x}_i + b = 0 \tag{1}$$

where w is the weight vector and b is biased. For the data of +1 and -1 classes, they meet the Inequality (2).

$$\begin{aligned} \underset{\sim}{w}^T . \underset{\sim}{x}_i + b &\leq -1 \\ \underset{\sim}{w}^T . \underset{\sim}{x}_i + b &\geq +1 \end{aligned} \tag{2}$$

The optimal hyperplane equation can be written in equation (3), the range of values α_i is $0 \leq \alpha_i \leq C$ (Piccialli & Sciandrone, 2022).

$$L = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n} \alpha_i \alpha_j \underset{\sim}{y}_i \underset{\sim}{y}_j \underset{\sim}{x}_i \underset{\sim}{x}_j \tag{3}$$

To solve non-linear problems, a kernel trick can be used. A polynomial kernel is one of the kernels that directs the classifier algorithm to a more flexible decision boundary, especially since the data is transformed into higher-dimensional spaces involving polynomials with degrees p. In general, the polynomial kernel function equation can be written with Equation (4).

$$K(\underset{\sim}{x}_i, \underset{\sim}{x}_j) = (\underset{\sim}{x}_i{}^T \underset{\sim}{x}_j + 1)^p \tag{4}$$

with p being a polynomial degree of value greater than $1 (p > 1)$. SVM basically has a linear principle, but it has been developed to overcome non-linear problems by incorporating the kernel concept into high-dimensional spaces, where optimal hyperplanes are sought to maximize margins between classes (Utari, 2023). SVM is inherently a linear classifier but has been extended to handle non-linear problems by applying kernel functions that map the input into higher-dimensional spaces in which an optimal hyperplane can be found (Virmani & Pandey, 2023).

b. Ensemble Bagging

Ensemble bagging (bootstrap aggregating) is one of the approaches in machine learning that is effective to improve the accuracy and stability of predictions, especially in dealing with the problem of imbalanced data (Gao et al., 2025). This technique works by generating a number of subsets of training data using bootstrap resampling, then training several basic models (weak learners) independently and in parallel on each subset (Breiman, 1996). The ensemble bagging algorithm enhances model generalization by reducing variance through the combination of multiple predictors trained on different resampling subsets, thereby increasing accuracy and stability in unbalanced classification problems (Arini et al., 2025). The ensemble bagging method is able to improve classification accuracy compared to the classic method, especially in data with unbalanced proportions, because the bootstrap aggregation process can reduce misclassification and improve model stability (Arini et al., 2025). The bagging work process can be seen in Figure 2, which is an adaptation of the illustrations by (Cendani & Wibowo, 2022).

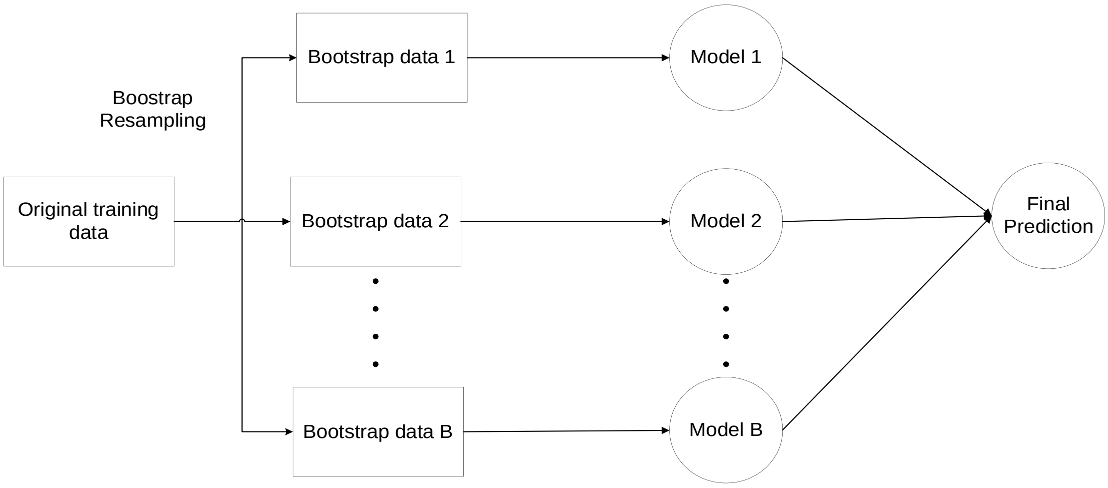

Figure 2. Work Process Ensemble Bagging

The procedure of the bagging algorithm can be summarized as follows.
1) Bootstrap Sampling
2) Independent Model Training wit SVM
3) Aggregate Prediction with Majority voting is carried out by Equation (5).

$$\widehat{y_B} = argmax_j f(x, \mathcal{L}_B) \tag{5}$$

where $\widehat{y_B}$ is the final prediction result based on the aggregation of all models at input x.

c. Classification Accuracy
A confusion matrix is used to find out how much data is correctly classified into the corresponding classes and how much of the incorrect data is classified into the incorrect classes (Migni et al., 2024). The following is a *confusion matrix* for the classification of two classes (Gao et al., 2025), as shown in Table 1.

Table 1. Confusion Matrix

Confusion Matriks		Prediction Class	
		Class +1	Class -1
Actual Class	Class +1	TP	FN
	Class -1	FP	TN

The following is the formula used in measuring the accuracy of classification.
1) Accuracy: Accuracy measures how well a classification model correctly predicts all classes as a whole that can be written in Equation (6).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

2) Sensitivity: Sensitivity or *true positive rate* measures how well the classification model identifies all true positive cases that can be written in Equation (7).

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

3) Specificity: Specificity or *true negative rate* measures the model's ability to correctly identify a sample that is actually negative that can be written in Equation (8).

$$Spesificity = \frac{TN}{FP + TN} \tag{8}$$

5. Steps

The research process began with collecting and describing respondents' socio-demographic data, followed by data preprocessing and standardization. The optimal SVM kernel function was then selected to develop the baseline model. Next, a Bagging ensemble was applied to improve SVM performance. Model accuracy, sensitivity, and specificity were evaluated, and the results were interpreted to highlight their implications for public health. A flowchart illustrating these steps is provided below to visually represent the methodological approach employed in this study, as shown in Figure 3.

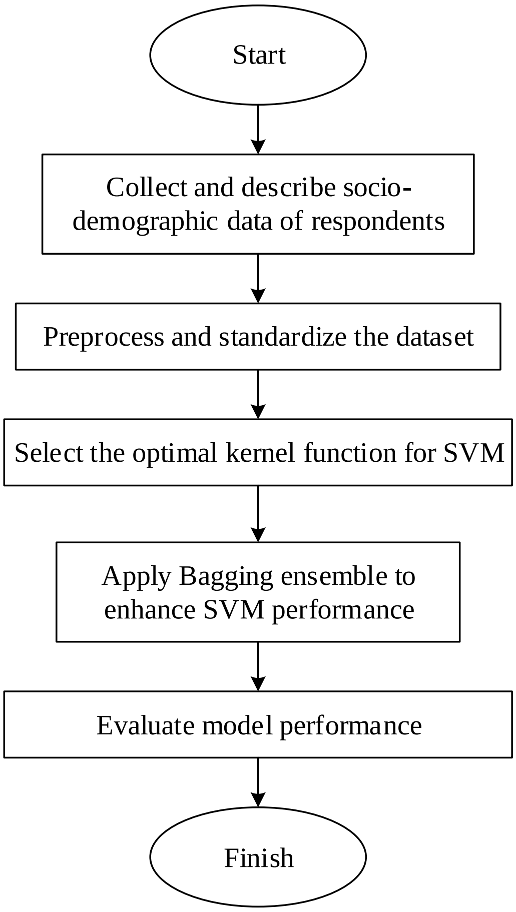

Figure 3. Research Workflow

C. RESULT AND DISCUSSION

1. Data Description

The characteristics of the respondents in this study were used to provide an overview of the socio-demographic background of mothers under five in Dadapan Village, Wajak District, Malang Regency. This data includes the mother's age, last level of education, and the family's economic status. These characteristics can affect the mother's knowledge, parenting patterns, and the fulfillment of child nutrition, which ultimately has an impact on the risk of stunting, as shown in Table 2.

Table 2. Respondent Characteristics

Respondent Characteristics	Category	Number (n=100)	Percentage (%)
Mother's Age (years)	< 25	18	18,0
	25–34	46	46,0
	35–44	27	27,0
	≥ 45	9	9,0
Final Education	No school	6	6,0
	Elementary/Equivalent	29	29,0
	Junior High School/Equivalent	32	32,0
	High School/Equivalent	25	25,0
	College	8	8,0
Status Economy	Low	44	44,0
	Intermediate	39	39,0
	Tall	17	17,0

Based on Table 3, the majority of respondents are in the age group of 25–34 years (46%), which is a productive period as well as the ideal age in childcare. Respondents with the age of <25 years (18%) showed that there were young mothers who had the potential to have limited experience in caring for toddlers. Meanwhile, respondents aged ≥45 years were relatively few (9%), reflecting the involvement of older mothers. In terms of education, most respondents completed junior high school (32%) and elementary school (29%), while 25% of those who reached high school completed high school, and only 8% received higher education. This condition illustrates that the level of education of mothers at the research site is still relatively low, so it has the potential to affect nutritional literacy and childcare patterns. The economic status of the family shows that almost half of the respondents are in the low category (44%), followed by the middle (39%) and high (17%) categories. This indicates that most families have limitations in purchasing power, especially related to meeting the nutritional needs of children under five. These socioeconomic conditions are a significant risk factor for stunting events.

In addition, the focus of the analysis is directed at the relationship between maternal knowledge, economic level, and nutritional status. These two predictor variables are considered to have an important role in differentiating maternal nutritional conditions, so a classification method is needed that is able to capture nonlinear patterns while providing accurate prediction results. One suitable approach is the Support Vector Machine (SVM) because of its ability to build optimal separator functions in high-dimensional spaces. However, to improve model stability and reduce the risk of misclassification due to data variations, an SVM ensemble bagging approach is also used that combines several bootstrap models. This approach is expected to provide more robust performance in separating different nutritional status groups, as shown in Figure 4.

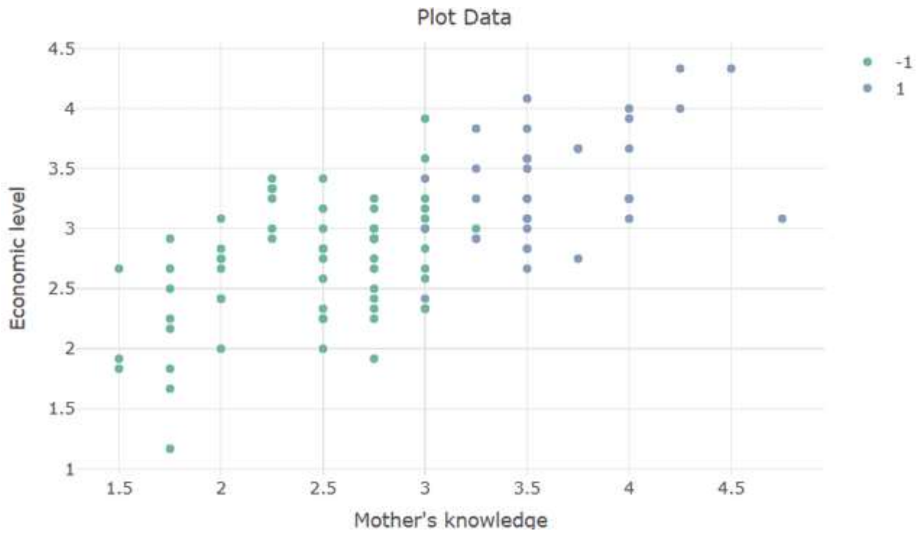

Figure 4. Plot Data Stunting

Based on Figure 4, it can be seen that the nutritional status of mothers can be differentiated according to two main variables, namely maternal knowledge and economic level. The group with low nutritional status (−1) tended to be in areas with low knowledge and economic level, while the group with good nutritional status (1) was dominant in areas with higher knowledge and economic level. Although the separation pattern is quite clear, there is overlap in the middle area, so some data points can potentially be misclassified when using only one SVM model. Conditions like this show that the SVM ensemble bagging method is very relevant to use, as it is able to reduce the variance of a single model by utilizing the prediction results of many bootstrap models.

2. Support Vector Machine (SVM)

The Support Vector Machine (SVM) produces *a hyperplane* or separator function that plays a role in classifying stunting risk. The model is built using a quadratic polynomial kernel with two predictor variables, namely X_1 (Mother Knowledge) and X_2 (Economic Level). The optimal hyperplane function obtained from the model training process is written as follows.

$$\hat{Y} = -12.353 + 0.231X_1 - 0.171X_2 + 1.02X_1^2 - 0.228X_2^2 + 0.336X_1X_2$$

The Support Vector Machine (SVM) model with a quadratic polynomial kernel is able to form an effective separator function (hyperplane) in classifying stunting risk based on maternal knowledge (X_1) and economic level (X_2). The resulting hyperplane equation shows that both variables, both individually and in quadratic interactions, contribute to classification. The positive coefficients at X_1 and X_1^2 confirmed that increased maternal knowledge tended to reduce the risk of stunting, while the negative coefficients at X_2 and X_2^2 indicated that there were certain limitations on the influence of economic levels. The positive interaction of 0.336

X_1X_2 also shows that the combination of maternal knowledge and economic level has an important role in differentiating risk categories.

3. Ensemble Bagging SVM

Ensemble Bagging SVM is a method that builds multiple SVM models from bootstrap data to improve predictive reliability. Each model was trained on a random sample of the training data, then the prediction results were combined with a majority voting technique. This approach is able to reduce the variance of a single model and result in a more stable and accurate classification of stunting risks. The prediction results of each resampling in the testing data as well as the final prediction obtained are shown in Table 5.

Table 5. Predicted Results with SVM Ensemble Bagging

Data Testing	Resampling 1	Resampling 2	Resampling 3	...	Resampling 100	Final Prediction
1	-1	-1	-1	...	-1	-1
2	1	1	1	...	1	1
3	-1	-1	-1	...	-1	-1
4	1	1	-1	...	1	1
5	-1	-1	-1	...	-1	-1

After obtaining the final prediction results from the SVM bagging process in the data testing, the next step is to compile a confusion matrix to evaluate the model's performance. The confusion matrix provides an overview of the amount of data that has been successfully classified correctly and that has experienced misclassification for each class. In this way, evaluation metrics such as accuracy, sensitivity, and specificity can also be calculated that reflect the model's performance in distinguishing between classes. Then the accuracy of the classification of each confusion matrix can be calculated by calculating the accuracy, sensitivity and specificity. The following are the results of the accuracy of the classification formed, as shown in Table 7.

Table 7. Accuracy of Classification of each Model

Model	Dataset	Accuracy	Sensitivity	Specificity
SVM	training	0.85	0.875	0.825
	testing	0.85	0.9	0.8
	overall	0.85	0.88	0.82
Ensemble Bagging SVM	training	0.9625	0.975	0.95
	testing	0.95	1	0.94
	overall	0.96	0.977	0.946

Table 7 showed that the basic Support Vector Machine (SVM) model was able to achieve an accuracy of 85%, with a sensitivity of 90% and a specificity of 80%. This means that SVM is quite good at identifying children who are really at risk of stunting (high sensitivity), although there are still weaknesses in distinguishing children who are not at risk (lower specificity). This is important because mistakes in detecting healthy children as stunting can create an unnecessary burden of intervention. However, when the SVM Ensemble Bagging approach was applied, the model's performance improved significantly with an accuracy of 95%, perfect sensitivity (100%), and 94% specificity. These results show that SVM Bagging is superior in producing stable and consistent predictions. Perfect sensitivity means that all children who are truly stunted are successfully detected, so the risk of false negatives (unidentified stunted children) can be avoided. This is crucial, because in the context of public health, failure to detect stunting early can result in delayed nutritional interventions.

In addition, the increase in specificity in SVM bagging up to 94% showed that this method was also able to reduce the misclassification of normal children as stunting (Kumari & Vinod, 2024). This improvement indicates that ensemble-based models effectively enhance predictive precision by minimizing false classifications in sensitive health datasets. Thus, nutrition interventions can be more targeted, allowing health resources and stunting prevention programs to be allocated more efficiently (Mahajan et al., 2023).

The enhanced performance of the SVM Ensemble Bagging method holds significant public health implications. With higher sensitivity and specificity, it can be used to accurately identify at-risk children at an earlier stage, ensuring that timely and effective interventions can be carried out. This is particularly crucial in regions with high rates of stunting, where early detection is key to preventing long-term developmental issues. Moreover, the increased accuracy in distinguishing between stunted and non-stunted children allows for better allocation of resources in public health programs, ensuring that interventions are directed to those most in need. By improving classification precision, this method could contribute to more effective and cost-efficient public health policies aimed at reducing stunting in Indonesia and other regions with similar challenges.

5. Disscussion

The application of the Ensemble Bagging Support Vector Machine (SVM) method in this study illustrates the integration of mathematical modelling and statistical learning in addressing classification problems in public health data. The SVM framework is formulated through quadratic optimization, aiming to determine an optimal separating hyperplane that maximizes the margin between two classes (stunting and non-stunting). The use of a quadratic polynomial kernel allows nonlinear transformation of the predictor space, enabling complex relationships between maternal knowledge and economic level to be represented in a mathematically tractable form. This transformation is fundamental in applied mathematics, where kernel mapping functions act as tools to linearize nonlinear decision boundaries. The resulting classification function represents a multivariate quadratic surface, combining main and interaction effects of the predictors. Statistically, the positive coefficients on maternal knowledge indicate its direct contribution in reducing the probability of stunting, while the negative coefficients on economic level reveal diminishing returns when economic resources reach a saturation point. The interaction term $0.336X_1X_2$ mathematically confirms that maternal knowledge and economic level are interdependent factors in influencing nutritional outcomes, reflecting the presence of interaction effects in multivariate classification models. These findings are consistent with recent evidence showing that ensemble and kernel-based SVM approaches effectively model nonlinear dependencies in health-related data, leading to improved interpretability and predictive accuracy (Liu et al., 2023).

The ensemble bagging procedure further strengthens the model by incorporating statistical resampling through the bootstrap method. Each bootstrap sample generates an independent SVM estimator, and the aggregation through majority voting reduces model variance and enhances stability. From a statistical standpoint, this approach represents a variance–bias trade-off solution, where combining multiple weak learners minimizes overfitting and improves generalization. This aligns with the results of Yuningsih et al. (2023)**,** who demonstrated that integrating bagging with SVM and synthetic resampling methods significantly enhances classification performance in imbalanced datasets. Similarly, previous work by Pristyanto & Zein (2023) also supported that bagging-based ensemble frameworks reduce variance and improve robustness when dealing with limited and imbalanced samples. The improved accuracy, sensitivity, and specificity demonstrate that ensemble-based kernel optimization can yield a statistically robust and mathematically consistent classifier for imbalanced data. Overall, this study highlights that the Bagging-SVM model serves not only as a computational algorithm but as a mathematical, statistical framework that bridges theory and application. The model effectively combines principles of optimization, kernel transformation, and resampling to produce a stable predictive system. Consistent with these results, ensemble learning continues to be recognized as an effective paradigm for health data classification, offering superior generalization performance compared to single classifiers (Yuningsih et al., 2023)**.** Such an approach demonstrates the potential of applied mathematics and statistics in supporting early detection systems and evidence-based public health policies, particularly in mitigating stunting risk through data-driven decision support.

D. CONCLUSION AND SUGGESTIONS

This study concludes that the Support Vector Machine (SVM) method with a quadratic polynomial kernel effectively classifies the risk of stunting based on maternal knowledge and family economic level. While the method showed good accuracy in classification, its performance improved significantly when combined with the Ensemble Bagging SVM approach. The Bagging SVM model achieved 95% accuracy, 100% sensitivity, and 94% specificity, demonstrating its robustness in detecting stunting cases without compromising precision in identifying healthy children. These findings confirm that Ensemble Bagging SVM is a powerful tool for early detection of stunting risk and a valuable addition to public health and nutrition intervention programs. This research contributes to the theoretical understanding of machine learning in public health, particularly in addressing imbalanced datasets in stunting risk classification. By integrating kernel transformations and ensemble bagging methods, the study bridges a gap in the application of non-linear classification techniques in health data analysis, an area often underexplored in existing literature. Methodologically, the study proposes a novel combination of SVM with polynomial kernels and ensemble bagging, providing a robust framework that can be used in similar classification problems in public health. The results also have practical implications, suggesting that machine learning models like SVM with Ensemble Bagging should be integrated into digital health systems for early detection of stunting. Such integration into mobile health applications or village health information systems would help health practitioners make more accurate and timely decisions, ultimately improving intervention efficiency.

From a policy perspective, the study suggests that public health policies aimed at reducing stunting should incorporate machine learning tools for more efficient and accurate early detection systems. The integration of Ensemble Bagging SVM into digital platforms can aid in real-time monitoring and ensure that interventions are targeted appropriately. However, technical challenges remain, including the need for adequate data infrastructure, data privacy concerns, and potential resistance to new technology from local health practitioners. Addressing these barriers will require collaboration between health authorities, technology developers, and local communities. While the study provides valuable insights, it is limited by the sample size of 100 respondents, which may not capture the full diversity of stunting cases across Indonesia. Future research should aim to expand the dataset to include more geographic regions and explore additional features such as maternal health history and socio-economic factors. Longitudinal studies could help assess the effectiveness of early detection tools over time, and future work could explore the integration of other machine learning techniques, such as deep learning or hybrid ensemble models, to further improve the accuracy and robustness of stunting risk prediction.

ACKNOWLEDGEMENT

The author would like to thank all parties who have contributed to the completion of this research. Special thanks are conveyed to the supervisor for the direction and input that is very meaningful in the process of preparing this scientific paper. Thank you also to colleagues and parties who have assisted in the collection of data and the implementation of the analysis. This research would not be completed properly without the support, motivation, and assistance of these various parties.

REFERENCES

Arini, L. H. Y., Solimun, Efendi, A., & Fernandes, A. A. R. (2025). Ensemble bagging with ordinal logistic regression to classify toddler nutritional status. *BAREKENG: Journal of Mathematics and Its Applications*, *19*(1), 1–12. https://doi.org/10.30598/barekengvol19iss1pp1-12

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/https://doi.org/10.1007/BF00058655

Cendani, L. M., & Wibowo, A. (2022). Comparison of ensemble learning methods in the classification of diabetes diseases. *Journal of the Informatics Society*, *13*(1), 33–44. https://doi.org/10.14710/jmasif.13.1.42912

Dongoran, R. Z., & Cipta, H. (2025). Robust optimization model analysis for online sentiment issues on Shopee using support vector machine. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *9*(3), 930–945. https://doi.org/10.31764/jtam.v9i3.31555

Gao, X., Xie, D., Zhang, Y., Wang, Z., Chen, C., He, C., ... & Zhang, W. (2026). A comprehensive survey on imbalanced data learning. *Frontiers of Computer Science, 20*(11), 2011622. https://arxiv.org/abs/2502.08960

Handayani, T., Sifriyani, S., & Dani, A. T. R. (2023). Nonparametric spline truncated regression with knot point selection method generalized cross validation and unbiased risk. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *7*(3), 848–863. https://doi.org/10.31764/jtam.v7i3.14034

Hattle, M., Burke, D. L., Trikalinos, T., Schmid, C. H., Chen, Y., Jackson, D., & Riley, R. D. (2022). Multivariate meta-analysis of multiple outcomes: characteristics and predictors of borrowing of strength from Cochrane reviews. *Systematic Reviews, 11*(1), 149. https://doi.org/10.1186/s13643-022-01999-0

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/https://doi.org/10.1109/TKDE.2008.239

Karthikeyan, S., & Ravichandran, K. S. (2022). SVM ensemble training for imbalanced data classification using multi-objective optimization techniques. *Applied Intelligence, 53*(12), 15424-15441.https://doi.org/10.1007/s10489-022-04291-9

Kumari, S. A. M., & Vinod, V. (2024). An Ensemble Machine Learning-Based Classification for Cardiovascular Disease Prediction Using PCA and SVM with Bagging. *Journal of Neonatal Surgery*, *14*, 4019. https://doi.org/10.52783/jns.v14.4019

Laksono, A. D., Sukoco, N. E. W., Rachmawati, T., & Wulandari, R. D. (2022). Factors Related to Stunting Incidence in Toddlers with Working Mothers in Indonesia. *International Journal of Environmental Research and Public Health*, *19*(17), 10654. https://doi.org/10.3390/ijerph191710654

Liu, Z., Li, Y., Chen, N., Wang, Q., Hooi, B., & He, B. (2023). Class-imbalanced learning on graphs: A survey. *ACM Computing Surveys, 57*(8), 1-16. https://doi.org/10.1145/3718734

Mahajan, R., Gupta, R., & Kumar, S. (2023). Performance Optimization of SVM-Based Ensemble Learning for Medical Data Classification Using Hybrid Bagging Approach. *Biomedical Signal Processing and Control*, *85*, 104871. https://doi.org/10.1016/j.bspc.2023.104871

Migni, A., Bartolini, D., Marcantonini, G., Tognoloni, A., Ceccarini, M. R., & Galli, F. (2024). Multivariate Data Analysis Methods and Their Application in Environmental Science. *Mass Spectrometry Reviews, 77*(4):e70068. https://doi.org/10.1002/mas.220017

Ministry, of H. of the R. of I. (2025). *SSGI 2024: Survey of Indonesia's Nutritional Status in Numbers*. Ministry of Health of the Republic of Indonesia. https://www.badankebijakan.kemkes.go.id/survei-status-gizi-indonesia-ssgi-2024/

Piccialli, V., & Sciandrone, M. (2022). Nonlinear optimization and support vector machines. *Annals of Operations Research*, *314*(1), 15–47. https://doi.org/10.1007/s10479-022-04655-x

Pristyanto, Y., & Zein, A. A. (2023). Model Balanced Bagging Berbasis Decision Tree Pada Dataset Imbalanced Class. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *12*(1), 9–15. https://doi.org/10.32736/sisfokom.v12i1.1399

Putri, P. M., Shafira, A. S., & Mahardhika, G. S. (2024). Stunting reduction strategy in Indonesia: Maternal knowledge aspects. *The Indonesian Journal of Public Health*, *19*(2), 329–343. https://doi.org/https://doi.org/10.20473/Ijph.v19i2.2024.329-343

Rumagit, G. A., & Pangkey, F. (2021). Bagging-based ensemble analysis in handling unbalanced data on classification modelling. *Klabat Accounting Review*, *2*(2), 165–178. https://doi.org/10.60090/kar.v2i2.589.165-178

Talakua, M. W., & Tomasouw, B. P. (2023). Design of KIP Kuliah selection system and recipient determination using Support Vector Machine (SVM). *BAREKENG: Journal of Mathematics and Its Applications*, *17*(3), 1803–1814. https://doi.org/10.30598/barekengvol17iss3pp1803-1814

UNICEF. (2023). *State of the world's children: For every child, nutrition*. UNICEF. https://www.unicef.org/reports/unicef-annual-report-2023

Utari, D. R. (2023). Application of the support vector machine (SVM) method in classification of hypertension. *BAREKENG: Journal of Mathematics and Its Applications*, *17*(4), 3523–3532. https://doi.org/https://doi.org/10.30598/barekengvol17iss4pp2263-2272

Victora, C. G., Christian, P., Vidaletti, L. P., Gatica-Domínguez, G., Menon, P., & Black, R. E. (2021). Revisiting maternal and child undernutrition in low-income and middle-income countries: Variable progress towards an unfinished agenda. *The Lancet*, *397*(10282), 1388–1399. https://doi.org/https://doi.org/10.1016/S0140-6736(21)00394-9

Virmani, D., & Pandey, H. (2023). Comparative analysis on effect of different SVM kernel functions for classification. In D. Gupta, A. Khanna, A. E. Hassanien, S. Anand, & A. Jaiswal (Eds.), *International Conference on Innovative Computing and Communications – Lecture Notes in Networks and Systems* (Vol. 492, pp. 657–670). Springer. https://doi.org/10.1007/978-981-19-3679-1_56

Wattimena, A. A., Tuakia, M. T., & Manuputty, A. P. (2023). Ensemble resampling Support Vector Machine, multinomial regression to multiclass imbalanced data. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(1), 269–280. https://doi.org/10.30598/barekengvol18iss1pp0269-0280

WHO (2020). *Nutrition landscape information system (NLiS) country profile indicators: Interpretation guide.* World Health Organization. https://www.who.int/about/accountability/results/who-results-report-2020-2021

Yuningsih, L., Pradipta, G. A., Hermawan, D., Ayu, P. D. W., Hostiadi, D. P., & Huizen, R. R. (2023). IRS-BAG-Integrated Radius-SMOTE Algorithm with Bagging Ensemble Learning Model for Imbalanced Data Set Classification. *Emerging Science Journal*, *7*(5), 1501–1516. https://doi.org/10.28991/ESJ-2023-07-05-04