**Zero : Jurnal Sains, Matematika, dan Terapan**
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 3, 2025
DOI: 10.30829/zero.v9i3.26818
Page: 781-793

# Ensemble Bagging Support Vector Machine–Kernel Discriminant Analysis Model for Stunting Potential Classification

[1] Alfiyah Hanun Nasywa
Department of Statistics, Universitas Brawijaya, Malang City, 65145, Indonesia

[2] Solimun
Department of Statistics, Universitas Brawijaya, Malang City, 65145, Indonesia

[3] Achmad Efendi
Department of Statistics, Universitas Brawijaya, Malang City, 65145, Indonesia

[4] Adji Achmad Rinaldo Fernandes
Department of Statistics, Universitas Brawijaya, Malang City, 65145, Indonesia

[5] Celia Sianipar
Department of Statistics, Universitas Brawijaya, Malang City, 65145, Indonesia

[6] Fachira Haneinanda Junianto
Department of Mathematics, Universitas Brawijaya, Malang City, 65145, Indonesia

**Article Info**

***Article history:***

Accepted, 26 December 2025

***Keywords:***

Classification;
Ensemble Bagging;
Kernel Discriminant Analysis;
Stunting;
Support Vector Machine.

**ABSTRACT**

Considering the maternal knowledge, economic status, and maternal nutritional status, the current study created an optimal risk assessment model to detect childhood stunting risk. At the same time, these variables are unbalanced and interrelated in non-linear fashion. Then, to these ends, an Ensemble Bagging model consisting of Support Vector Machine and Kernel Discriminant was trained by voting on the aggregation of the majority of 100 bootstrapped samples, which countered variance and overfitting reducing, hence improving generalization. The primary data were sourced from the mothers of toddlers in the Wajak District. The model predicators were 3 out of the primary ones accounting for the stunting risk. The model also recorded an accuracy of 95%, sensitivity level of 80%, as well as a 100% specificity score. Non-linear relationships were detected and the variance was also reduced, supporting the study to place itself in the realms of novelty by being the first research to fuse the Ensemble Bagging with Kernel methods for Detected stunting risk, The model, hence, fits best as a decision Support System for detecting stunting risk at an early stage.



***Corresponding Author:***

Solimun,
Department of Statistics,
Universitas Brawijaya, Malang City, Indonesia
Email: solimun@ub.ac.id

## 1. INTRODUCTION

The focus of this technique is to perform the classification of instances as Algorithmic Pattern Recognition is centered on how observations can be assigned to different categories based on their characteristics. Classification attempts to separate the different classes by means of response variables (predictive variables) which delineate the regions of the response classes to make predictions on new observations [1]. With the advancement of computer science, other forms of classification like Decision Trees, Naïve Bayes SVM, KNN, CART, and Discriminant Analysis have surfaced, each taking different heuristics to map points of the classification problem.

Support Vector Machine (SVM) is one of the most popular classification algorithms. SVM aims to separate instances by finding the hyperplane that optimizes the margin between the classes [2]. SVM is very efficient when the data points are of high dimension and when it is also non-linear, this is mitigated by the kernel tricks that are employed like linear, polynomial, and radial base (RBF) [3]. On the other hand, SVM tends to underperform when the data set is imbalanced. This is because the model will be more biased to the majority set of classes. Thus, additional methods are needed to enhance the model performance.

Discriminant Analysis also serves as a robust method for partitioning a dataset according to the values of certain predictive variables by forming a weighted average of said predictive variables and maximizing the differences among classes [1]. Two primary forms of the method exist, which are Lineal Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), each forming their own predictions based on the assumption of equal covariance among classes. But LDA and QDA can be suboptimal when one is faced with linearly separable data [4].

To address the identified problem, KDA (Kernel Discriminant Analysis) is one of the extensions of the method which employs a technique commonly used by Support Vector Machines (SVM) [5]. As a result of the kernel technique, the data is mappable into higher orders of dimensionality where a linear separation is guaranteed [6]. This technique enhances the performance of predictive models on data containing a high degree of complexity and on which classical methods of discrimination are suboptimal. The predictive performance of models classification can be further elevated when using this technique with methods of ensemble learning, particularly bagging [7]. This method improves versatility when applied to data that is unstable or is imbalanced.

The lagging support related to appropriate methods to tackle stunting as a nutrition-related chronic problem highlights the need to customize methods to support classifying such ailments to chronic malnutrition stunting as low height for age [7]. SSGI 2024 from the Ministry of Health of the Republic of Indonesia recorded a national stunting rate of 19.8%, while stunting in Malang Regency reached 16.4%, and the 2020-2024 RPJMN target of 14% has not been achieved [8]. Stunting in Wajak District, which is in the prioritized area for the acceleration of stunting, is particularly significant at the regional level.

There are other studies that suggest the use of polynomial kernels in SVM and KDA to improve accuracy, especiallyif the data follows some pattern of nonlinearity. To be more specific, in the work of [9] it was shown that polynomial kernel SVM gives the highest accuracy to the classification of data pertaining to COVID-19 and bears complex patterns. The work in [10] confirmed that the SVM polynomial kernel classification of Tuberculosis was also the best at 96.27% accuracy, outperforming the other kernels and Naïve Bayes. Same goes with [11] where he was able to show that Polynomial Kernel Discriminant Analysis was able to sustain above 82 % accuracy on the given data being of high-dimensional, albeit the data being on the original was quite high. In another hand, the studies on ensemble bagging as stated also showed good results on improving the accuracy and stability of the model as well. [12] demonstrated that bagging-SVM can improve the model accuracy to 95 % on the economic inequality data. [13] demonstrated that bagging on LDA and QDA improved the model stability on unbalanced data significantly.

The SVM, KDA, polynomial kernels, and bagging have been the subjects of multiple, but individual studies within the literature. While considering the complementary capabilities of SVM, and KDA, none have focused on the fusion of SVM and KDA into an ensemble bagging framework. This underrepresentation in the literature suggests a lack of a comprehensive strategy aiming to use both SVM, and KDA, to enhance the predictability of highly imbalanced datasets. This lack of predictability within datasets suggests an equally imbalanced class dataset, which is specifically evident within the datasets associated with the classification of stunting. The datasets associated with stunting classification present a highly imbalanced dataset with a variety of nonlinear datasets. The focus of this study is to propose an ensemble bagging methodology that incorporates SVM, and KDA with the objective of combining the strengths of each to overcome the classification of stunting and associated risk factors.

The goal of this study is to provide a nonlinear ensemble bagging model, which is an integration of SVM, and KDA to determine the stunting associated risk. The use of SVM to maneuver the hyperplanes and KDA to funnel the data through an assortment of nonlinear features, this model will funnel the data to a singular point with a higher clarity than the models alone. This is the ultimate focus of this study is derive the empirical performance of the models to provide a boost in the system with higher data accuracy, improved sensitivity, and better system robustness. The efficiency of this system will close the gap of the system to provide a better data centered supportive decision system and aligned decision theories to formulate a stunting reduction program.

## 2. RESEARCH METHOD

This chapter describes the methodologies and the steps undertaken to devise the Ensemble Bagging Support Vector Machine - Kernel Discriminant Analysis System to classify (stunting potential). These activities include data collection, variable operationalization, and data preprocessing, followed by the construction of a model using kernel transformation and bagging and finally evaluating the model using metrics of accuracy, sensitivity, and specificity. A brief description of the dataset used in this research is provided below, followed by a detailed discussion of each methodological component.

### 2.1. Research Data

For the study, survey data was collected from 100 respondents. The sample consisted of mothers of children under 5 years old who reside in Dadapan Village, Wajak District, Malang Regency. The sample size can be considered valid given the complex multivariate analysis based on SVM and discriminant analysis as the study only has a limited number of predictor variables. Previous studies showed that SVM was suitable for smaller data sets, as long as the data was well structured and the number of dimensions were limited. Acknowledging the small sample size, future studies can include a wider population to enhance the result validity of the study. The samples were selected following the quota sampling technique and a questionnaire was administered based to obtain data from respondents based on a Likert scale of five. The variables under the study include Maternal Knowledge (X1), Economic Status (X2), Maternal Nutritional Status (X3), and the dependent variable, which was Stunting Potential (Y). The dependent variable is binary in scale (stunting and non-stunting), and is classified and modeled using the ensemble bagging technique. All the data for the study were processed, the model was trained, and the data were evaluated using RStudio for analysis, which made the analysis reproducible, and the statistics were uniform. The following R packages were used to aid in analysis: readxl to read data from Excel files, psych for reliability analysis, dplyr for data manipulation, and e1071 for building Support Vector Machine (SVM) models. Also, caret was used for cross-validation and measuring the performance of the models, while openxlsx was used to save the results.

To protect the data, strict inclusion and exclusion criteria were set. Only mothers of children aged 0 to 59 months and living in Dadapan Village, Wajak District, Malang Regency, and who agreed to participate by signing the informed consent form, were included in the study. Also, only those who were able to understand the instructions on how to fill out the questionnaire were included. The exclusion criteria included mothers of children under five years of age with an extreme congenital disease or abnormality that could impede their growth, and respondents who were not present at the research site during data collection or those who provided incomplete or invalid questionnaires. These criteria were set to make sure that the selected respondents are part of the target population, to obtain more accurate results that could be scientifically validated for the classification of stunting risk.

### 2.2. Determinants of Stunting

According to [8], stunting is defined as a condition of growth retardation in children starting from the first 1,000 days of pregnancy to the first 1,000 days of age, which is up to 23 months, due to chronic malnutrition. This definition is strengthened by [14] and anthropometric data using indicators of Body Length to Age (PB/U) or Height to Age (TB/U), with a Z-score value below $-2$ standard deviation as the limit of stunting children. Nutritional status assessment and stunting identification based on Z-score calculation can be done with equation (1).

$$Z = \frac{Observed\ Child\ Height - Mean\ Reference\ Growth}{Standard\ Deviation\ of\ Reference\ Growth} \tag{1}$$

The average value and standard deviation refer to the national anthropometric standards as stated in the Regulation of [15]. The following are the limits of the critical point of Z-score stunting.

**Table 1.** Height Categories based on Z-score

| Z-Score Critical Points | Category |
|---|---|
| X ≤ -2SD | Stunting |
| -2SD < X < 2SD | Usual |
| X ≥ 2SD | Tall |

Several different variables can affect the prevalence of stunting in children in East Java including: socio-economic status, level of education, employment status, the age of the children, as well as the differences between urban and rural environments [16]. Stunting in children under the age of five occurs as a result of chronic malnutrition, and is defined as being a height that is lower than the age standard [17]. The role of mothers in childcare is very important because mothers spend more time with their children than fathers, so they have a deeper understanding of the conditions of child growth and development [18] Mothers also play a role in regulating children's diet and physical activity, which can create a supportive environment for physical and mental health. The level of maternal nutrition knowledge has been proven to affect children's consumption patterns. [19] stated that mothers with good nutritional knowledge are able to provide nutritious food according to the needs of toddlers. found that mothers who did not get comprehensive information related to the concept of cascading had a lower level of nutritional knowledge than mothers who obtained information through social media or posyandu cadres. Cascading explains the flow-on effect of nutritional problems not attended to, as malnutrition during one phase of development leads to further complications at subsequent stages.

Besides the knowledge factor, families' economic condition is also an essential determinant on the fulfillment of toddlers' nutrition. Indonesia is still experiencing a potential middle income trap, which is stagnation of economic growth with subsequent implications of low purchasing power of the populace [21]. Families of low economic condition also tend to have low access to food of high nutritional value, thus the nutritional demands of the children are not sufficiently met. [22] mentions low income families as those with a high vulnerability of experiencing many nutritional problems such as low weight at birth, underweight, and stunting. Maternal nutritional status also is a cardinal factor on the risk of children stunting. There is a high probability of stunting among children born to mothers with a height of less than 150 cm, stunting is as a result of growth intergenerational limitation [23]. This condition is often the result of a long standing malnutrition history of the mother during teenage years. Also, how healthy a pregnant woman is can determine how healthy her baby is going to be. If a pregnant woman is not healthy and well-nourished, her baby could be born with growth problems [24]. Nutritional deficiencies during pregnancy can lead to intrauterine growth disorders, which then increases the risk of stunting in the early stages of life. In addition, eating habits and family lifestyle since pregnancy also affect the nutritional condition of the baby after birth, so attention to the nutritional status of the mother is a key step in comprehensive stunting prevention efforts.

### 2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the techniques used to achieve classification by attempting to identify the optimal linear hyperplane that will act as the decision boundary to divide observations that belong to different classes [2]. To compliment this technique further is the linear classification of cases that can also be extended via the use of kernel functions, leading to better classification results as opposed to those offered by traditional methods [25]. When kernel functions are utilized, it is possible to achieve non-linear mapping of data such that the data of two classes can be separated by a hyperplane that exists in a space of sufficiently high dimensionality [26].

There is likely a non-linear relationship concerning maternal knowledge, economic, and nutritional status with the risk of stunting, and the kernel trick of SVM affords the ability to model such complex relations. Furthermore, SVM is more effective when working with high-dimensional spaces, which is necessary since the predictors in this case involve multiple dimensions of maternal health and socio-economic status. Besides, SVM is especially a good fit when the data is numerical, which is exactly the case when it comes to the variables of this study, maternal knowledge scores, economic status, and nutritional status indicators. The numerical data enables SVM to effectively map the data points into higher dimensions for the purpose of improved classification which is more accurate and robust to perturbations [27].

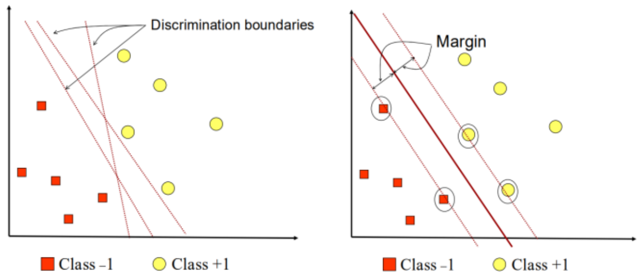


**Figure 1.** Hyperplane that separates the two classes

Suppose each data is denoted as $(\underset{\sim}{x}_i, \underset{\sim}{y}_i)$ , where the vector $\underset{\sim}{x}_i$ variable is a predictor and $\underset{\sim}{y}_i \in \{-1, +1\}$ is a class label [2]. The two classes are said to be perfectly separable if there is a hyperplane with an equation (2).

$$\underset{\sim}{w}^T . \underset{\sim}{x}_i + b = 0 \tag{2}$$

where w is the weight vector and b is biased. For the data of +1 and -1 classes, they meet the inequality (3).

$$\underset{\sim}{w}^T . \underset{\sim}{x}_i + b \leq -1 \tag{3}$$

$$\underset{\sim}{w}^T . \underset{\sim}{x}_i + b \geq +1$$

In data that cannot be separated linearly from the data, soft margin SVM is used by adding $\xi_i > 0$ slack variables and C penalty parameters to control the trade-off between margin width and misclassification. The misclassification controlling parameter C is the same for the margin's tradeoff of width. When the margin is entirely closed, misclassifications are heavily penalized with large values of C. Alternatively, small values of C give more tolerance to the misclassification of data, and widen the margin. Whilst C is above quoted margin control, it maintains its core function. The equation can be written in equation (4).

$$\min_{w,b,\xi} \frac{1}{2}\|\underset{\sim}{w}\|^2 + C\sum_{i=1}^{n}\xi_i \tag{4}$$

Provided:

$$\underset{\sim}{y}_i(\underset{\sim}{w}.\underset{\sim}{x}_i + b) \geq 1 - \xi_i \;, \xi_i \geq 0 \tag{5}$$

Through the Lagrange approach, this optimization is transformed into a form of dual problem that can be written in equation (6) in the range of values $\alpha_i$ is $0 \leq \alpha_i \leq C$.

$$L = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\alpha_i\alpha_j\underset{\sim}{y}_i\underset{\sim}{y}_j\underset{\sim}{x}_i\underset{\sim}{x}_j \tag{6}$$

A kernel trick is one means to approach non-linear problems. One such kernel is the polynomial kernel which leads the polynomial classifier algorithm to a more flexible decision boundary because the data has been transformed into polynomial functions of higher dimensions. The polynomial kernel function equation is given in equation (7).

$$K(\underset{\sim}{x}_i, \underset{\sim}{x}_j) = (\underset{\sim}{x}_i{}^T\underset{\sim}{x}_j + 1)^p \tag{7}$$

with $p$ being a polynomial degree of value greater than 1 $(p > 1)$.

Here, $p$ is representing which polynomial will be chosen and has a direct impact on which polynomial will be chosen along with how complex it will be. In this study, the chosen $p$ = 2 is a result which is the product of hyperparameter tuning. This result shows that it is a quadratic kernel which models the best flexibility along with generalization performance. Model flexibility 2 is the sweet spot for this cause as it allows the classifier to capture non-linear interaction effects without excessive model complexity and overfitting which is exactly what is required for the stunting classification data of this research. SVM is based on linear principles, but has evolved to deal with non-linear cases using the high dimensional kernel concept to find optimal hyperplanes that maximize the margin between classes [28].

## 2.4. Kernel Discriminant Analysis (KDA)

Kernel Discriminant Analysis (KDA) is a newer branch of Linear Discriminant Analysis (LDA) developed to analyze data when class dividing boundaries are nonlinear [24]. In contrast to LDAs that look for the best linear projection in the original space, KDA maps data to higher-dimensional feature spaces through φ(x) nonlinear transformations, so that linear separation between classes can be performed in those feature spaces [5]. KDA fits the characteristics of the stunting dataset in this research better than LDA. The predictor variables (maternal knowledge, economic level, and maternal nutritional status) have non-linear interactions that LDA will skip due to its linearity restriction. Preliminary analysis indicated that LDA resulted in overlapping class boundaries while KDA clearly separated the potentially stunting and non-stunting classes after the kernel transformation. This proves that KDA is better able to model different characteristics of the data non-linearly. In the feature space, the scatter matrix is calculated for the class φ(x) and against the class. The scatter matrix is defined as in Equation (8) in the class of the feature space and in Equation (9) for the scatter matrix of the classes in the feature space.

$$\mathbf{S}_\mathbf{B}^{\boldsymbol{\Phi}} = \sum\nolimits_{k=1}^{g} n_k(\overline{\underset{\sim}{x}}_k^{\phi} - \overline{\underset{\sim}{x}}^{\phi})(\overline{\underset{\sim}{x}}_k^{\phi} - \overline{\underset{\sim}{x}}^{\phi})' \tag{8}$$

$$\mathbf{S}_\mathbf{W}^{\phi} = \sum_{k=1}^{g}(n_k - 1)\mathbf{S}_k = \sum_{k=1}^{g}\sum_{j=1}^{n_k}(\phi(\underset{\sim}{x}_{kj}) - \overline{\underset{\sim}{x}}_k^{\phi})(\phi(\underset{\sim}{x}_{kj}) - \overline{\underset{\sim}{x}}_k{}^{\phi})' \tag{9}$$

where $\underset{\sim}{\overline{x}}_k^{\phi}$ is the J-class average in the feature space and $\underset{\sim}{\overline{x}}^{\phi}$ is the global average in the feature space. Linear discriminators can be maximized by equation (10).

$$J(\underset{\sim}{\widehat{w}})\frac{\underset{\sim}{\widehat{w}}'\mathbf{S}_{\mathbf{B}}^{\boldsymbol{\phi}}\underset{\sim}{\widehat{w}}}{\underset{\sim}{\widehat{w}}'\mathbf{S}_{\mathbf{W}}^{\phi}\underset{\sim}{\widehat{w}}}=\frac{\underset{\sim}{\widehat{w}}'\sum_{k=1}^{g}n_k(\underset{\sim}{\overline{x}}_k^{\phi}-\underset{\sim}{\overline{x}}^{\phi})(\underset{\sim}{\overline{x}}_k^{\phi}-\underset{\sim}{\overline{x}}^{\phi})'\underset{\sim}{\widehat{w}}}{\underset{\sim}{\widehat{w}}'\left[\sum_{k=1}^{g}\sum_{j=1}^{n_k}(\phi(\underset{\sim}{x}_{kj})-\underset{\sim}{\overline{x}}_k^{\phi})(\phi(\underset{\sim}{x}_{kj})-\underset{\sim}{\overline{x}}_k^{\ \phi})'\right]\underset{\sim}{\widehat{w}}} \tag{10}$$

with $\underset{\sim}{w} \in F$.

The Fisher-based kernel discriminant approach is obtained by integrating the kernel functions into the equation (10) and applying the extension form $\underset{\sim}{w}$ of the vector as shown in Equation (11).

$$\underset{\sim}{w}=\sum_{j=1}^{n_k}\alpha_j\phi(x_j) \tag{11}$$

The equation in the kernel discriminant analysis based on the Fisher approach is obtained by maximizing the function in the following equation (12).

$$J(\alpha)=\frac{\alpha' M\alpha}{\alpha' N\alpha} \tag{12}$$

where M and N are a matrix formed from the kernel values between samples.
After obtaining α from Fisher optimization, new data x is projected using the kernel function in equation (13).

$$\left(\underset{\sim}{w}\phi(x)\right)=\sum_{j=1}^{n}\alpha K(\underset{\sim}{x}_j,\underset{\sim}{x}) \tag{13}$$

## 2.5. Ensemble Bagging

Ensemble bagging (bootstrap aggregating) is one of the approaches in machine learning that is helpful in improving the accuracy and the stability of the prediction especially in the case of imbalanced data [2]. Ensemble bagging (bootstrap aggregating) is one of the approaches in machine learning that is helpful in improving the accuracy and the stability of the prediction especially in the case of imbalanced data [7]. The bagging ensemble method improves the accuracy of the classification of results compared to the classical method especially in the case of imbalanced data because the bootstrap aggregation process improves stability and reduces misclassification of the data [29]. The work process of bagging can be seen in Figure 2.4 which is an adaptation of the illustrations of [30].

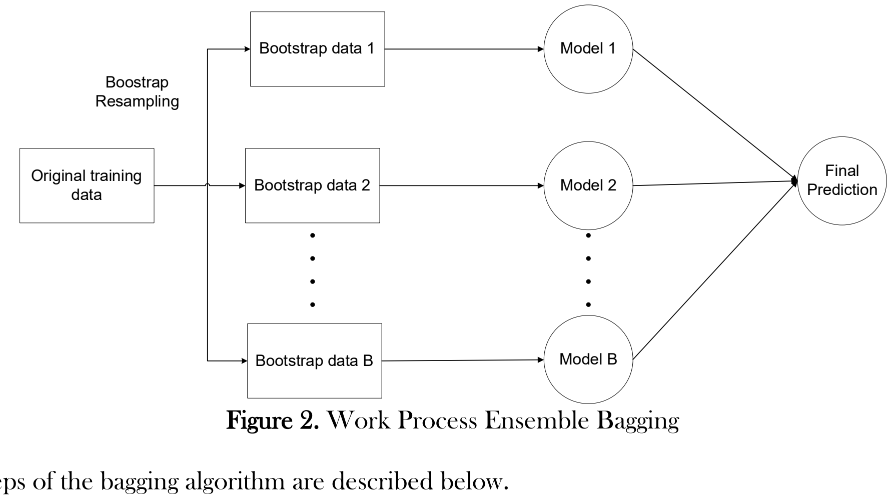


**Figure 2.** Work Process Ensemble Bagging

The steps of the bagging algorithm are described below.

1) Bootstrap Sampling
   Conduct n Bootstraps of the L training sample data set. Each sample is drawn uniformly at random with replacement, yielding an LB bootstrap sample.
2) Independent Model Training
   The subsets obtained by resampling are used to train the predictive modeling algorithm independently. In this case, the bagging algorithm was used with SVM (Support Vector Machine) as the predictor.

3) Agregasi Prediksi (Aggregate Prediction)
When all the models finish training, their predictions are combined to arrive at a final outcome. In the case of classification, this is done using the majority voting method; for regression, the output is the mean of all predictions. The majority voting is done in accordance with equation (14).

$$\widehat{y_B} = argmax_j f(x, \mathcal{L}_B) \tag{14}$$

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

### 2.6. Classification Accuracy

One of the purposes of a confusion matrix is to assess how much data is correctly assigned to the appropriate classes and how much of the wrong information is assigned to the incorrect classes. Below is a confusion matrix for a case of two-class classification [2].

**Table 2.** Confusion Matrix

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

Information:
TP : The quantity of data for which the predicted value class and the true value class are the same
FP : The quantity of data for which the predicted value class is +1 and the true value class is -1
TN : The quantity of data for which the predicted value class and the true value class are the same
FN : The quantity of data for which the predicted value class is -1 and the true value class is +1

The following is the formula used in measuring the accuracy of classification.

a. Accuracy: Accuracy measures how well a classification model correctly predicts all classes as a whole that can be written in Equation (15).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{15}$$

b. Sensitivity: Sensitivity or *true positive rate* measures how well the classification model identifies all true positive cases that can be written in Equation (16).

$$Sensitivity = \frac{TP}{TP + FN} \tag{16}$$

c. 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 (17).

$$Specificity = \frac{TN}{FP + TN} \tag{17}$$

### 2.7. Research Steps

This sequential stage of research has been designed to guide the SVM-KDA Ensemble Bagging model development process, so that it is able to reliably predict the class of stunting potential. The research is in the following stages:

1) Performs bootstrapping to generate 100 training data resamples.
2) For each of these resamples, two models are trained: SVM with a quadratic polynomial kernel, and KDA with a quadratic polynomial kernel.
3) On the test data, SVM and KDA all generate predictions.
4) For the final predictions, the class is decided by majority vote.
5) For both training and test data, a confusion matrix is created.
6) The measures of accuracy, sensitivity, and specificity are calculated.
7) The performance of the model is explained in the context of the class imbalance.

# 3. RESULT AND ANALYSIS

## 3.1. Data Description

This study involved 100 mothers who had children under five in Dadapan Village, Wajak District, Malang Regency. The characteristics of the respondents were analyzed based on three main variables, namely Maternal Knowledge ($X_1$), Economic Level ($X_2$), and Maternal Nutritional Status ($X_3$), as well as the Potential for Stunting (Y) variable. A descriptive summary of each variable is presented in Table 3.

**Table 3.** Variable Characteristics

| Variable | Indicator | Average | |
|---|---|---|---|
| | | Indicator | Variable |
| Mother's knowledge ($\boldsymbol{X_1}$) | The term stunting ($X_{1.1}$) | 2,685 | 2,868 |
| | Nutritional knowledge ($X_{1.2}$) | 3,050 | |
| Economy level ($\boldsymbol{X_2}$) | Family income ($X_{2.1}$) | 2,795 | 2,949 |
| | Family expenses ($X_{2.2}$) | 3,103 | |
| Nutritional status of the mother ($\boldsymbol{X_3}$) | Type of food ($X_{3.1}$) | 3,360 | 3,372 |
| | Number of meals ($X_{3.2}$) | 3,383 | |

The said variables were also described overall, and were categorized as medium, in the case of Mother's Knowledge, Economic Level as well as Maternal Nutritional Status were all in the medium category. Mother's Knowledge had in average, 2.868 on the understanding of stunting and child nutrition. While the Economic Status of Respondents had 2.949, and that is relatively moderate for the family's Economic Status. Maternal Nutritional Status had an average of 3.372 which means that the mother's food intake or diet is generally enough. All in all, these means that these families are relatively in good condition however, the family's knowledge, economy and nutrition are of great importance in order to reduce the odds of stunting on children, and therefore must the family condition improve in those aspects in order to improve the stunting odds the children in the family are facing. In addition to the above descriptive characteristics, an assessment of the relationships in Mother's Knowledge, Economic Level, and Maternal Nutritional Status in relation to the potential stunting, had also been done through the use of interval data with Scatter Plots for the purpose of visualization.

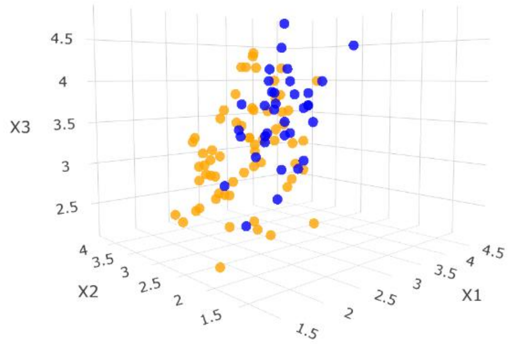


**Figure 3.** Scatter Plots Between Variables

Figure 3 shows that there is no clear pattern of separation along the potentially stunting and non-stunted classes, as the dots for the two classes overlap each other in data space a lot. This condition indicates that a simple linear combination of the three independent variables fails to identify the relation to stunting potential, so that the hyperplanes that separate the classes cannot be determined by linear lines or planes alone. The unusual distributions provide evidence of the nonlinear relationship, particularly in the regions where maternal knowledge and nutritional status intersect. As a result, kernel-based methods such as SVM and KDA have become much more applicable, as they are capable of relocating data to higher-dimensional spaces and capturing more complex geometric arrangements of data to enhance the precision of classification of stunting potential.

## 3.2. Optimal Kernel Determination

The choice of the kernel used in SVM is determined to be the best mapping function to separate the classes in a nonlinear fashion. A comparison is done for the three kernels in use which includes linear kernel, polynomial kernel, and radial kernel in Figure 4.

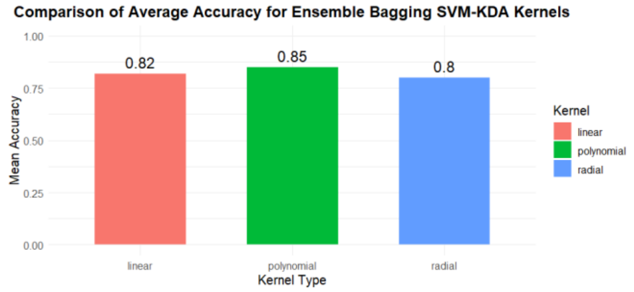

**Figure 4.** Kernel Optimal

The evaluations revealed that with a score of 0.82, the linear kernel performed the best, with the radial kernel following behind at 0.80, and the polynomial kernel having the highest score of 0.85. The polynomial kernel's superiority suggests that they are better at identifying and capturing nonlinear relationships and patterns in the specified data. Thus, the polynomial kernel was chosen as the best kernel in the SVM–KDA Bagging Ensemble modeling.

### 3.3. Model Ensemble Bagging SVM-KDA

When deploying the SVM-KDA Ensemble Banding and SVM and KDA models, 100 resampling bootstrap replicates were taken to fit the SVM and KDA models. Each resampling model is different based on the random samples drawn from the data. Below exhibits the 1st to the 100th resampling models and the marginal coefficients from SVM and KDA models.

It is agreed that the number of 100 bootstrap iterations is the optimal number, as it is thought to be the number that is enough to stabilize ensemble-based performance measure while not being too computationally expensive. Preliminary tests on the model show that after 80 to 100 iterations of resampling, the model's accuracy, sensitivity, and specificity plateau. Hence, 100 bootstrap iterations is the optimal number as it balances computational efficiency,

SVM and KDA models in the 1st resampling

$$\hat{Y}_{SVM} = 2{,}004 + 0{,}055X_1 - 0{,}240X_2 - 0{,}647X_3 - 0{,}481X_1^2 + 0{,}691X_2^2 + 0{,}251X_3^2 - 0{,}213X_1X_2 + 0{,}145X_1X_3 - 0{,}385X_2X_3$$

$$\hat{Y}_{KDA} = -0{,}270 + 0{,}011X_1 + 0{,}004X_2 - 0{,}031X_3 - 0{,}027X_1^2 - 0{,}084X_2^2 - 0{,}236X_3^2 - 0{,}149X_1X_2 + 0{,}632X_1X_3 + 0{,}591X_2X_3$$

SVM and KDA models in the 2nd resampling

$$\hat{Y}_{SVM} = 1{,}340 + 2{,}131X_1 + 1{,}542X_2 + 0{,}574X_3 - 0{,}126X_1^2 + 0{,}892X_2^2 + 0{,}074X_3^2 - 0{,}706X_1X_2 + 0{,}476X_1X_3 - 0{,}472X_2X_3$$

$$\hat{Y}_{KDA} = -0{,}268 + 0{,}008X_1 - 0{,}019X_2 - 0{,}026X_3 + 0{,}106X_1^2 + 0{,}099X_2^2 - 0{,}206X_3^2 + 0{,}400X_1X_2 - 0{,}356X_1X_3 + 0{,}131X_2X_3$$

Using this model and the technique of majority voting, we achieved the results of the prediction. This technique reduces the variance from one model and results in a more consistent and precise classification of the risk of potential stunting. The prediction outputs of every resampling in the test data along with the last prediction attained are presented in the following table.

**Table 4.** 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 |
| ... | ... | ... | ... | ... | ... | ... |
| 20 | -1 | -1 | -1 | ... | -1 | -1 |

After getting the last forecast outcomes from the SVM-KDA bagging procedure while testing the data, the following stage is to create a confusion matrix to analyze the model's performance. A confusion matrix displays a summary of the total data for each class that has been appropriately classified and that has been (re)classified incorrectly. Thus, performance evaluation metrics like accuracy, sensitivity, and specificity, can be determined which measure the model's class separation effectiveness. Table 6 displays the results of the confusion matrix for SVM bagging in testing the data.

**Table 5.** Confusion Matrix Ensemble Bagging SVM

| | | | |
|---|---|---|---|
| **Data Training** | | | |
| Confusion Matrix | | Prediction Class | |
| | | Class 1 | Class -1 |
| Actual Class | Class 1 | 46 | 8 |
| | Class -1 | 4 | 22 |
| **Data Testing** | | | |
| Confusion Matrix | | Prediction Class | |
| | | Class 1 | Class -1 |
| Actual Class | Class 1 | 4 | 1 |
| | Class -1 | 0 | 15 |

Then the classification of the confusion matrix can yield accuracy and can yield accuracy by calculating the sensitivity and specificity. What follows are the results of the classification accuracy.

**Table 6.** Accuracy of Classification of each Model

| Dataset | Accuracy | Sensitivity | Specificity |
|---|---|---|---|
| Training | 0,850 | 0,733 | 0,920 |
| Testing | 0,950 | 0,800 | 1 |

Referring to Table 7, the SVM-KDA Ensemble Bagging model demonstrated remarkable classification ability for both datasets. On the training data, the model achieved an accuracy of 0.850 with a sensitivity of 0.733 and specificity of 0.920 which indicates the model can determine a majority of stunting cases while also being highly accurate in determining non-stunting cases. On the testing data, the performance improved with an accuracy of 0.950 and a sensitivity of 0.800 and a specificity of 1.000. These outcomes show that the model not only excelled in determining cases on the training data, but also demonstrated continued proficiency and stability in the classification of the new data, even able to pinpoint all non-stunting cases with no mistakes.

These findings suggest that the SVM-KDA Bagging Ensemble shows excellent generalization and can be trusted for the early detection of the risk of stunting. The fusion of SVM and KDA bagging has helped bolster the resilience of the model in the presence of data imbalance and nonlinear relationships among indicators. The model's perfect specificity on the test data attests to its efficacy in steering clear of false positives in the normal population, while the boost in sensitivity indicates that the model has the potential to identify toddlers who are genuinely at risk. Thus, this method ought to be employed at the community and primary health care level for stunting risk prevention.

### 3.4. Disscussion

These findings indicate the practical and methodological implications for risk assessment and prediction of stunting incidence in children under five. In these findings, the SVM-KDA Ensemble Bagging model provided improvements in accuracy and measures of true prediction in sensitivity and specificity relative to single models. These results align with findings that support the ensemble approach as an effective means of variance reduction and increased stability on non-homogeneous and complex data for classification [31]. The combination of SVM hyperplane and the KDA discriminant function is advantageous because both can capture complex dynamics and non-linear relationships that have been emphasized in literature which deal with kernel-class modern learning and the mapping to high-dimensional spaces [32]. Furthermore, this study adds to the body of knowledge that demonstrates the effectiveness of bootstrap-based resampling techniques in addressing class imbalances which is common in public health data [2].

Given the circumstances, the outcome of the model has some possibilities that can be employed in the decision support system of the stunting reduction acceleration program. Extreme specificity indicates that the model can reduce the false positive rate in the normal subgroup of preschool toddlers in which case the intervention can be directed towards those that are genuinely in need. This strategy evaluates the evidence that has shown that systems of machine learning-based risk detection are able to elevate the impact of intervention on public health challenges through efficient detection of the risk case [33]. Conversely, the model sensitivity indeed reflects that the cases of potential stunting are even more accurately identified in which case proactive intervention can be introduced to avoid the chronicity of the problem in terms of nutrition.

The other implications have to do with sociological and family parameters. The economic position and knowledge of a mother and the mothers' nutritional status are key variables of stunting risk based on the findings of this study, which has been documented in other studies advocating the need for family-based approaches to strengthen the nutritional behavioral patterns and the ability to acquire diversified foods [34]; [16]. Hence, for instance, the AI-derived classification models such as the study conducted could be useful for educational program planning, nutrition risk classification and intervention targeting at a given village or health center. This strategy also aims at the recent studies advocating for the application of machine intelligence in health management systems [35].

In summation, the present study demonstrates the integration of SVM–KDA with ensemble bagging contributes not only to the health classification body of knowledge and methodology, but also the real-world application to stunting reduction paradigm shift in Indonesia. The model has the potential to be advanced as a mobile application or as a risk tracking system that health personnel can utilize to enhance effort and precise targeting for the identification of children under five years with stunting.

## 4. CONCLUSION

This research shows that the SVM-KDA-based Ensemble Bagging method enhances the detection and classification of stunting risks compared to the performance of a sole model. This approach, using SVM hyperplane and KDA discriminants, models the class boundaries more effectively on nonlinear data. The rest of the prediction model retains bagging, which is stabilizing. All measures of performance increased, including accuracy and generalization, as the model performed better at the boundaries on unseen data and generalization. This is measurable across all parameters of the training and testing sets. This research confirms that kernel-based machine learning techniques can be significant as a decision support mechanism for the early diagnosis of stunting.

It has been established that the created model has value that extends beyond the specific methodology. Such a model can also be applied to the public health monitoring system. However, every study has its bounds, and this one is no exception. Because the model used a single region, the data quality and data representativeness can limit the model's predictive accuracy. For this reason, future studies should contain empirical data from several regions to improve the model's versatility. Additionally, the forecasts might improve if more exogenous factors, such as/including the environment or genes, were considered. As such, this study lays the groundwork for innovative data-determined measures of risk that could be used to devise more effective strategies aimed to swiftly reduce the levels of stunting in Indonesia.

## ACKNOWLEDGEMENT

The author gratefully acknowledges all those who have helped during the course of this research. The author especially thanks the supervisor for his/her major guidance and input throughout the author's preparation of this scientific paper. Also, thanks to the colleagues and people who have helped the author in collecting and analyzing the data. This research would not have been successfully completed without the encouragement and help of so many people.

## 5. REFERENCES


[1] A. Migni, D. Bartolini, G. Marcantonini, A. Tognoloni, M. R. Ceccarini, and F. Galli, "Multivariate Data Analysis Methods and Their Application in Environmental Science," *Mass Spectrom. Rev.*, 2024, doi: 10.1002/mas.220017.

[2] X. Gao *et al.*, "A Comprehensive Survey on Imbalanced Data Learning," *arXiv Prepr.*, vol. arXiv:2502, 2025, [Online]. Available: https://arxiv.org/abs/2502.08960

[3] J. Shao, X. Liu, and W. He, "Kernel Based Data-Adaptive Support Vector Machines for Multi-Class Classification," *Mathematics*, vol. 9, no. 9, p. 936, 2021, doi: 10.3390/math9090936.

[4] R. Ran, T. Wang, Z. Li, and et al., "Polynomial linear discriminant analysis," *J. Supercomput.*, vol. 80, pp. 413–434, 2024, doi: 10.1007/s11227-023-05485-9.

[5] Y. Jin, X. Zhang, and A. J. Molstad, "Kernelized Discriminant Analysis for Joint Modeling of Multivariate Categorical Responses," *J. Comput. Graph. Stat.*, pp. 1–12, 2025, doi: 10.1080/10618600.2025.2526412.

[6] L. Qu and Y. Pei, "A Comprehensive Review on Discriminant Analysis for Addressing Challenges of Class-Level Limitations, Small Sample Size, and Robustness," *Processes*, vol. 12, no. 7, p. 1382, 2024, doi: 10.3390/pr12071382.

[7] P. M. Putri, A. S. Shafira, and G. S. Mahardhika, "Stunting reduction strategy in Indonesia: Maternal knowledge aspects," *Indones. J. Public Heal.*, vol. 19, no. 2, pp. 329–343, 2024, doi: https://doi.org/10.20473/Ijph.v19i2.2024.329-343.

[8] of H. of the R. of I. Ministry, *SSGI 2024: Survey of Indonesia's Nutritional Status in Numbers*. Jakarta: Ministry of Health of the Republic of Indonesia, 2025.

[9] F. Prasetiyowati, N. Fitriyah, A. Setiawan, and A. S. Nugroho, "Comparative Analysis of Machine Learning Algorithms for COVID-19 Case Classification Using Polynomial Kernel SVM," *IEEE, IEEE.* doi: 10.1109/ICICoS57375.2023.10132269.

[10] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.isssue3.year2020.page133-143.

[11] A. Araveeporn and A. Kangtunyakarn, "An Enhanced Discriminant Analysis Approach for Multi-Classification with Integrated Machine Learning-Based Missing Data Imputation," *Mathematics*, vol. 13, no. 21, p. 3392, 2025, doi: 10.3390/math13213392.

[12] H. A. Shiddiqi, K. E. Setiawan, and R. Fredyan, "Leveraging Support Vector Machines and ensemble learning for early diabetes risk assessment: A comparative study," *J. EMACS (Engineering, Math. Comput. Sci.*, vol. 7, no. 1, pp. 1–6, 2025, doi: 10.21512/emacsjournal.v7i1.12846.

[13] Solimun and A. A. R. Fernandes, "Ensemble bagging discriminant and logistic regression in classification analysis," *New Math. Nat. Comput.*, vol. 21, no. 1, pp. 91–111, 2025, doi: 10.1142/S1793005725500061.

[14] W. H. O. (WHO), *Nutrition landscape information system (NLiS) country profile indicators: Interpretation guide*. World Health Organization, 2020.

[15] Pemenkes, *Regulation of the Minister of Health of the Republic of Indonesia Number 2 of 2020 concerning Indonesian Child Anthropometric Standards*. Ministry of Health of the Republic of Indonesia, 2020.

[16] A. D. Laksono, N. E. W. Sukoco, T. Rachmawati, and R. D. Wulandari, "Factors Related to Stunting Incidence in Toddlers with Working Mothers in Indonesia," *Int. J. Environ. Res. Public Health*, vol. 19, no. 17, p. 10654, 2022, doi: 10.3390/ijerph191710654.

[17] L. H. Y. Arini, Solimun, A. Efendi, and A. A. R. Fernandes, "Ensemble bagging with ordinal logistic regression to classify toddler nutritional status," *BAREKENG J. Math. Its Appl.*, vol. 19, no. 1, pp. 1–12, 2025, doi: 10.30598/barekengvol19iss1pp1-12.

[18] Y. B. Prasetyo, P. Permatasari, and H. D. Susanti, "The effect of mothers' nutritional education and knowledge on children's nutritional status: a systematic review," *Int. J. Child Care Educ. Policy*, vol. 17, p. 11, 2023, doi: 10.1186/s40723-023-00114-7.

[19] M. H. A. Syahroni, N. Astuti, V. Indrawati, and R. Ismawati, "Factors that affect the eating habits of preschool-age children (4–6 years) are reviewed from the achievement of balanced nutrition," *J. Culin. Arts*, vol. 10, no. 1, pp. 12–22, 2021, doi: https://ejournal.unesa.ac.id/index.php/jurnal-tata-boga/.

[20] K. Patiran, T. Siswati, and E. Yuliati, "Relationship between Maternal Knowledge and Nutritional Status of Children in Teluk Patipi, Fakfak," *JPK J. Prot. Kesehat.*, vol. 13, no. 1, pp. 72–76, 2024, doi: 10.36929/jpk.v13i1.816.

[21] H. W. Falah and Syafri, "Determination of Economic Growth in Indonesia," *J. Econ. Trisakti*, vol. 3, no. 2, pp. 2309–2318, 2023, doi: https://doi.org/10.25105/jet.v3i2.16541.

[22] H. Rohmawati, N. L. M. Puspita, A. Awatiszahro, and A. Nugroho, "The Relationship Between Family Economic Level and the Incidence of Anemia in Pregnant Women," *Str. J. Ilm. Kesehat.*, vol. 13, no. 1, pp. 31–37, 2024, doi: 10.30994/sjik.v13i1.1107.

[23] I. P. Sari, Y. Ardillah, and A. Rahmiwati, “Berat bayi lahir dan kejadian stunting pada anak usia 6-59 bulan di Kecamatan Seberang Ulu I Palembang,” *J. Gizi Indones.*, vol. 8, no. 2, pp. 110–118, 2020, doi: 10.14710/jgi.8.2.110-118.
[24] R. D. Aisyah, S. Suparni, and F. Fitriyani, “Effect of Counseling Packages on The Diet of Pregnant Women With Chronic Energy Deficiency,” *Str. J. Ilm. Kesehat.*, vol. 9, no. 2, pp. 944–949, 2020, doi: 10.30994/sjik.v9i2.399.
[25] M. W. Talakua and B. P. Tomasouw, “Design of KIP Kuliah selection system and recipient determination using Support Vector Machine (SVM),” *BAREKENG J. Math. Its Appl.*, vol. 17, no. 3, pp. 1803–1814, 2023, doi: 10.30598/barekengvol17iss3pp1803-1814.
[26] A. S. Nugroho, D. Handoko, and A. B. Witarto, *Support Vector Machine – Theory and Its Application in Bioinformatics*. 2003.
[27] D. Mustafa Abdullah and A. Mohsin Abdulazeez, “Machine Learning Applications based on SVM Classification A Review,” *Qubahan Acad. J.*, vol. 1, no. 2, pp. 81–90, 2021, doi: 10.48161/qaj.v1n2a50.
[28] D. R. Utari, “Application of the support vector machine (SVM) method in classification of hypertension,” *BAREKENG J. Math. Its Appl.*, vol. 17, no. 4, pp. 3523–3532, 2023, doi: https://doi.org/10.30598/barekengvol17iss4pp2263-2272.
[29] L. H. Y. Arini, A. Solimun, A. Efendi, and M. O. Ullah, “CART classification on ordinal scale data with unbalanced proportions using ensemble bagging approach,” *JTAM (Jurnal Teor. dan Apl. Mat.*, vol. 8, no. 2, pp. 441–453, 2024, doi: 10.31764/jtam.v8i2.20201.
[30] L. M. Cendani and A. Wibowo, “Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes,” *J. Masy. Inform.*, vol. 13, no. 1, pp. 33–44, 2022, doi: 10.14710/jmasif.13.1.42912.
[31] R. Qiu, “Func-Bagging: An Ensemble Learning Strategy for Imbalanced Classification,” *Appl. Sci.*, vol. 15, no. 2, p. 905, 2025, doi: https://doi.org/10.3390/app15020905.
[32] Z. Liu, Y. Li, N. Chen, Q. Wang, B. Hooi, and B. He, “Class-Imbalanced Learning on Graphs: A Survey,” *ACM Comput. Surv.*, 2023, doi: 10.1145/3718734.
[33] B. Rao, M. Rashid, M. G. Hasan, and G. Thunga, “Machine Learning in Predicting Child Malnutrition: A Meta-Analysis of Demographic and Health Surveys Data,” *Int. J. Environ. Res. Public Health*, vol. 22, no. 3, p. 449, 2025, doi: 10.3390/ijerph22030449.
[34] R. Aristiyani and M. Mustajab, “The relationship between family economic level and the incidence of stunting in toddlers,” *J. Nutr. Public Heal.*, vol. 12, no. 1, pp. 45–53, 2023.
[35] J. Amann *et al.*, “To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems,” *PLOS Digit. Heal.*, vol. 1, no. 2, p. e0000016, 2022, doi: 10.1371/journal.pdig.0000016.