

## RESEARCH ARTICLE

# Grey Wolf Optimization-Based Feature Selection for Improved Stunting Malnutrition Classification

Santosa <sup>1,2,\*</sup> and Zailani Abdullah <sup>3</sup>

<sup>1</sup>Department of Informatics, Faculty of Engineering, Universitas Muhammadiyah Maluku Utara, Ternate, Indonesia

<sup>2</sup>Faculty of Business, Management and Information Technology, Universiti Muhammadiyah Malaysia, Perlis, Malaysia

<sup>3</sup>Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Kota Bharu, Malaysia

\*Correspondence [p52400010@student.umam.edu.my](mailto:p52400010@student.umam.edu.my)

Received: 10 September 2026; Revised: 22 September 2026; Accepted: 29 September 2026; Published: 30 September 2026

## ABSTRACT

Stunting remains a child health problem that requires accurate identification of nutritional conditions from demographic and anthropometric data. Existing machine learning studies on stunting classification use different models and evaluation procedures, while the combined effect of class balancing and metaheuristic feature selection across multiple classifiers remains insufficiently examined. This study integrates Grey Wolf Optimization (GWO), Synthetic Minority Oversampling Technique (SMOTE), and ten machine learning classifiers for stunting malnutrition classification. The dataset contains 114 records of children under five years of age collected from two Community Health Centers in Semarang City, Indonesia, with 13 predictor attributes and one target variable. GWO was used to select feature subsets, while SMOTE was applied to address class imbalance. The evaluated classifiers were RF, DT, XGBoost, KNN, GBDT, AdaBoost, SVC, CatBoost, LightGBM, and Stacking, using accuracy, precision, recall, and F1-score as evaluation metrics. GBDT achieved the highest accuracy, precision, and F1-score among the evaluated models. Using all attributes, GBDT achieved 62.86% accuracy before SMOTE and 73.68% after SMOTE. After GWO-based feature selection, the accuracy increased to 82.86% before SMOTE and 86.84% after SMOTE. Recall showed a different pattern across the evaluated conditions. These findings show that GWO-based feature selection and SMOTE affect classification performance differently across evaluation metrics and provide a comparative framework for stunting analysis in health informatics.

## KEYWORDS

Stunting; Grey Wolf Optimization; Feature Selection; SMOTE; Machine Learning; Classification; Health Informatics

## 1. Introduction

Stunting is a condition of impaired linear growth caused by prolonged nutritional inadequacy and insufficient care during childhood [1]. Stunting affects children's physical growth and increases the risk of morbidity and mortality [2]. Children with stunting can also experience impaired cognitive development, reduced future

productivity, and long-term health problems [3]. These effects make stunting an important public health problem that requires systematic prevention and early identification.

Stunting remains a health problem in developing countries, including Indonesia [4]. Inadequate nutritional intake during critical periods of child development contributes to impaired height growth relative to age [5]. Health records from Semarang City also document stunting cases among children in several age groups [6]. Figure 1 presents the distribution of stunting cases in Semarang City according to age. These data show the need for a systematic method to analyze child growth information and identify nutritional conditions.

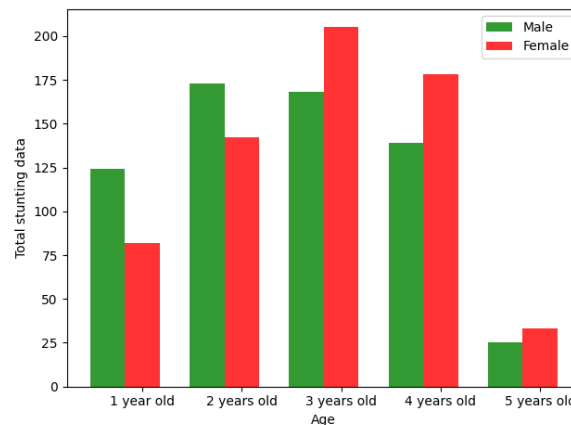


**Figure 1.** Distribution of stunting cases in Semarang City according to age [6].

Machine learning (ML) provides computational methods for analyzing patterns in child health data. ML algorithms can process demographic characteristics, anthropometric measurements, environmental factors, and other indicators related to nutritional status [7]. Anthropometric indicators, including weight-for-age and height-for-age  $z$ -scores, provide quantitative information about child growth. Height-for-age is specifically used to assess linear growth and identify stunting conditions. ML models can use these variables to classify nutritional conditions based on patterns contained in the available data [8].

The performance of an ML model depends on the relevance of the features used during the learning process. Redundant and irrelevant features can increase model complexity without improving classification performance [9]. This problem becomes more important when the number of observations is limited relative to the number of candidate features. Feature selection addresses this problem by identifying a subset of features that contains relevant information for classification [10]. A reduced feature set also allows the classifier to focus on variables that directly contribute to class discrimination [11].

Feature selection can be formulated as an optimization problem because the algorithm must identify an appropriate subset from several possible feature combinations. The number of possible combinations increases as the number of available features increases. Exhaustive evaluation therefore requires the examination of many candidate subsets. Metaheuristic optimization provides a search mechanism for identifying suitable feature subsets without evaluating every possible combination. Grey Wolf Optimization (GWO) is a swarm intelligence algorithm that has been applied to feature-selection problems [12].

GWO models the social hierarchy and hunting behavior of grey wolves. The algorithm divides candidate solutions into four hierarchical groups: alpha, beta, delta, and omega. The alpha represents the best current solution, while the beta and delta represent the second and third best solutions [13]. The remaining search agents update their positions according to these three leading solutions. This mechanism controls the exploration of the search space and the exploitation of promising solutions. In feature selection, each search agent represents a candidate feature subset, while the fitness function evaluates the quality of that subset.

Feature selection does not address class imbalance in a dataset. Class imbalance occurs when one outcome class contains more observations than another class. This condition can influence a classifier because the learning process receives more information from the majority class. As a result, overall accuracy alone does not provide a complete description of classification performance. Precision, recall, and F1-score are therefore required to evaluate model performance from different perspectives.

Synthetic Minority Oversampling Technique (SMOTE) addresses class imbalance by generating synthetic observations for the minority class [14]. SMOTE generates these observations from the relationship between minority samples and their nearest neighbors [14]. This process increases the representation of the minority

class in the training data without directly duplicating existing observations. The balanced training data provide classifiers with a more proportional representation of both classes. The effect of SMOTE must be evaluated separately for each classification model because different algorithms use different learning mechanisms.

Previous studies have applied ML algorithms to stunting and malnutrition classification. Other studies have examined feature-selection techniques and class-balancing methods for classification problems. However, three methodological issues remain relevant to the present study. First, different classifiers can produce different responses to class imbalance because each algorithm applies a different decision mechanism. Second, feature selection can produce different feature subsets and classification results when it is combined with different classifiers. Third, separate evaluations of feature selection and class balancing do not explain their combined effect on classification performance. These issues require a comparative framework that evaluates feature selection and class balancing under consistent experimental conditions.

This study addresses these issues by integrating GWO-based feature selection, SMOTE, and multiple machine learning classifiers for stunting malnutrition classification. GWO identifies feature subsets from the available stunting attributes. SMOTE balances the class distribution in the training data. The study evaluates ten classification models: Support Vector Classifier (GWO+SVC), K-Nearest Neighbors (GWO+KNN), Decision Tree (GWO+DT), Gradient Boosting Decision Trees (GWO+GBDT), Random Forest (GWO+RF), XGBoost (GWO+XGBoost), LightGBM (GWO+LightGBM), AdaBoost (GWO+AdaBoost), CatBoost (GWO+CatBoost), and Stacking Classifier (GWO+Stacking). The evaluation uses accuracy, precision, recall, and F1-score to compare model performance under different feature-selection and class-balancing conditions.

This study provides three contributions. First, the study integrates GWO-based feature selection with ten machine learning classifiers in a common experimental framework for stunting malnutrition classification. Second, the study evaluates each classifier before and after SMOTE under conditions with and without GWO-based feature selection. Third, the study compares the classification results using accuracy, precision, recall, and F1-score. This comparison provides empirical information about the effect of feature selection and class balancing on different machine learning models for the analyzed stunting dataset.

## 2. Related Works

Several studies have applied machine learning methods to childhood malnutrition and stunting classification. A hybrid ensemble model based on majority voting was developed to predict malnutrition among children under five years of age in sub-Saharan African countries. The model combined several machine learning algorithms and achieved an accuracy of 96%. The study also compared the ensemble model with Random Forest, Decision Tree, XGBoost, and K-Nearest Neighbors [15]. Machine learning classifiers were also evaluated for stunting classification among children under five years of age in Zambia. The study used the 2018 Zambia Demographic and Health Survey dataset and evaluated Logistic Regression, Random Forest, Support Vector Classifier, XGBoost, and Naive Bayes. Model performance was evaluated using accuracy, recall, precision, and F1-score. Random Forest achieved a testing accuracy of 79%, which was the highest value among the evaluated classifiers [16].

A Decision Tree model was developed to classify nutritional conditions among infants aged 0–59 months. The model classified stunting, wasting, and underweight conditions. The model used the Gini index and achieved an accuracy of 82.22% for stunting, 72.23% for wasting, and 78.35% for underweight. These results show that the Decision Tree model classified several forms of childhood malnutrition in the analyzed dataset [17]. Machine learning was also applied to predict nutritional requirements related to the growth of children and adolescents. The study used child-growth variables as inputs for predictive modeling. The results provide an example of machine learning application for nutritional assessment in children and adolescents [18]. Machine learning methods were also applied to malnutrition prediction among children under five years of age in India. The study analyzed health and demographic variables related to childhood malnutrition. The evaluated models classified nutritional conditions from the available child-health data. The study provides an example of machine learning application to population-based malnutrition data [19].

Machine learning methods were also evaluated for childhood stunting analysis in Bangladesh. The study compared gradient boosting, Random Forest, Support Vector Classifier, classification trees, and Logistic Regression. Gradient boosting produced the highest predictive performance among the evaluated models. The analysis identified child age, household wealth, maternal education, previous birth interval, paternal education, geographical area, household size, maternal age at first birth, maternal nutritional status, and parental age as variables associated with stunting. These findings show that model selection and variable selection influence childhood stunting analysis [20]. Longitudinal child-growth data were also analyzed using

machine learning methods. The study examined anthropometric and inflammatory indicators associated with child growth from birth to 48 months of age. Random Forest identified height-for-age  $z$ -score at birth as an important predictor of subsequent growth. The analysis also identified alpha-1-acid glycoprotein, C-reactive protein, and interleukin-1 as relevant inflammatory biomarkers. The study shows that machine learning can analyze anthropometric and biological variables within the same predictive framework [21].

The reviewed studies used different machine learning algorithms, datasets, and evaluation procedures for childhood malnutrition and stunting analysis. The studies examined individual classifiers, ensemble methods, and variable-selection approaches under different experimental settings. However, the reviewed studies did not evaluate class balancing and GWO-based feature selection across multiple classifiers within the same experimental framework. This study therefore evaluates SMOTE and GWO-based feature selection across ten machine learning classifiers under consistent experimental conditions.

### 3. Material and Method

This study developed a classification framework for stunting malnutrition using Grey Wolf Optimization (GWO), Synthetic Minority Oversampling Technique (SMOTE), and ten machine learning classifiers. The experimental process includes data preparation, class balancing, feature selection, model training, and performance evaluation. Figure 2 presents the workflow of the proposed method.

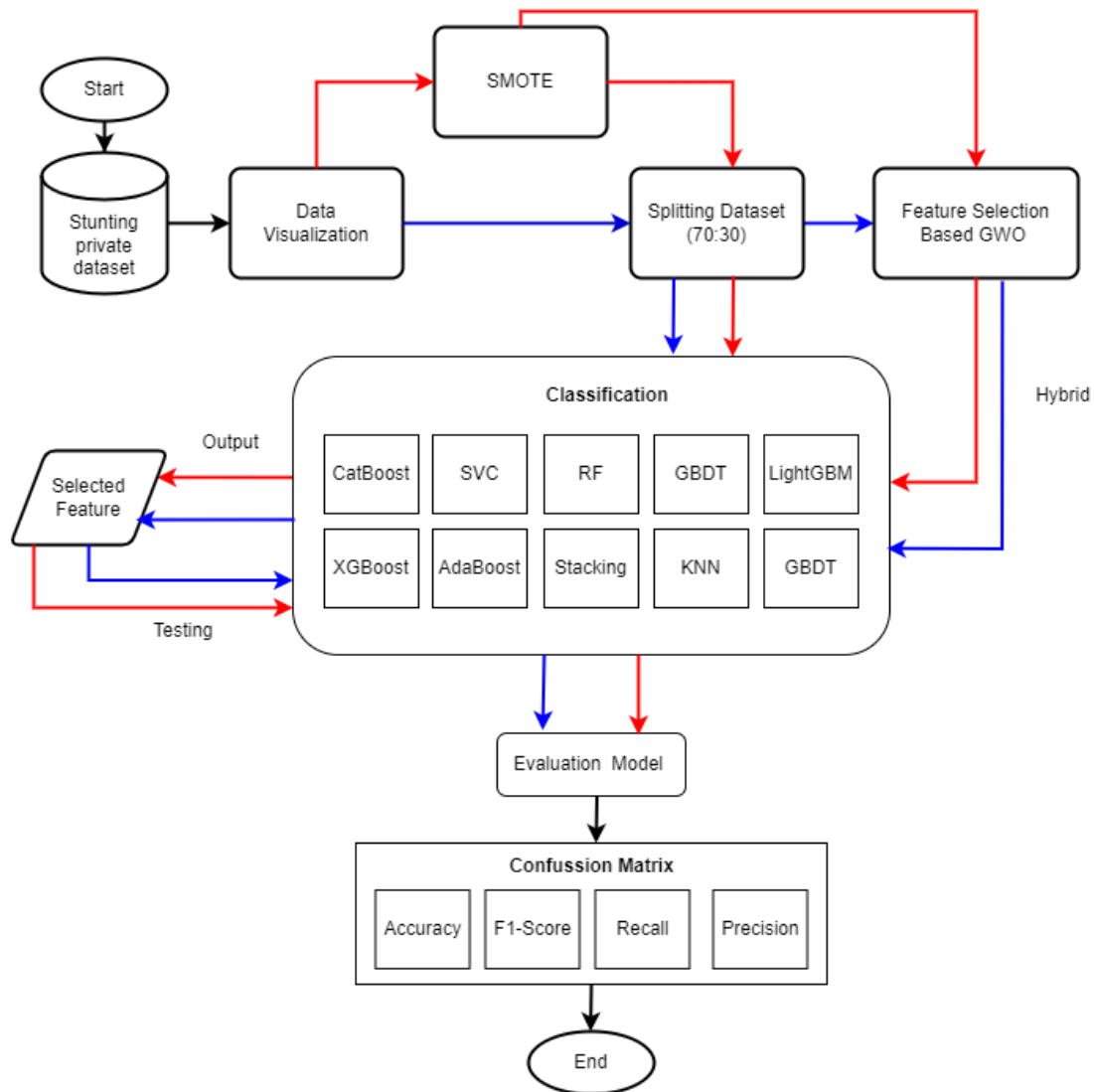


Figure 2. Workflow of the proposed method.

### 3.1. Data Source

The dataset was obtained from two Community Health Centers located in Kedungmundu and Tambakaji, Semarang City, Central Java, Indonesia. The dataset contains 114 records of children under five years of age. The recorded variables include demographic and anthropometric information related to nutritional status. Table 1 presents the variables, descriptions, and data types used in the dataset.

**Table 1.** Description of the stunting malnutrition dataset.

| Attribute    | Description                  | Data Type |
|--------------|------------------------------|-----------|
| Gender       | Sex of the child             | Integer   |
| Age          | Age of the child             | Integer   |
| Birth Weight | Birth weight of the child    | Float     |
| Birth Height | Birth height of the child    | Float     |
| Weight       | Current weight of the child  | Float     |
| Height       | Current height of the child  | Float     |
| LiLA         | Upper arm circumference      | Float     |
| W/A          | Weight-for-age indicator     | Integer   |
| ZS W/A       | Weight-for-age $z$ -score    | Float     |
| H/A          | Height-for-age indicator     | Integer   |
| ZS H/A       | Height-for-age $z$ -score    | Float     |
| Weight Gain  | Weight-gain indicator        | Integer   |
| ZS W/H       | Weight-for-height $z$ -score | Float     |
| Category     | Nutritional status           | Integer   |

### 3.2. Data Visualization

Data visualization was used to examine the distribution, relationships, and patterns of the variables in the stunting dataset. Graphical representations were used to summarize the characteristics of the data and support the identification of relevant patterns. The visualization process also supported the examination of variable distributions and relationships before model development. Different visualization forms were selected according to the characteristics of the analyzed variables and the objectives of the analysis.

### 3.3. Synthetic Minority Oversampling Technique

#### 3.4. Synthetic Minority Oversampling Technique

Synthetic Minority Oversampling Technique (SMOTE) was used to address class imbalance in the stunting dataset. SMOTE generates synthetic samples for the minority class based on the nearest neighbors of existing minority instances. The method creates new samples between selected minority instances and their neighboring samples instead of directly duplicating existing observations. This process increases the representation of the minority class and produces a more balanced class distribution for model training. The balanced training data were then used in the subsequent classification process [22].

### 3.5. Train-Test Split

The dataset was divided into training and testing sets using a 70:30 ratio. Seventy percent of the data was used for model training, while the remaining 30% was used for model testing. A random state of 42 was applied to maintain a consistent and reproducible data split across the experiments.

### 3.6. Model Classification

#### 3.6.1 Random Forest

Random Forest (RF) is an ensemble learning method that combines multiple decision trees for classification. Each tree is constructed using bootstrap samples and a randomly selected subset of features. The final class is determined from the predictions generated by the individual trees. For probabilistic classification, the final prediction is obtained by averaging the prediction probabilities produced by the trees. This mechanism allows

RF to combine information from multiple decision trees within a single classification model [23, 24]. The prediction of an RF model containing  $M$  trees can be expressed as

$$m_{M,n}(x, \theta_1, \dots, \theta_M, \vartheta_n) = \frac{1}{M} \sum_{m=1}^M m_n(x, \theta_m, \vartheta_n), \quad (1)$$

where  $M$  represents the number of trees,  $x$  denotes the input sample,  $\theta_m$  represents the random parameters associated with the  $m$ -th tree, and  $\vartheta_n$  represents the training data. Equation 1 calculates the RF output by averaging the predictions produced by the  $M$  individual trees.

### 3.6.2 Decision Tree

Decision Tree (DT) is a supervised learning model that performs classification through a hierarchical structure of decision rules. Each internal node evaluates a specific feature, each branch represents the outcome of the decision rule, and each leaf node represents the predicted class. In this study, DT uses the available child-related attributes to separate the data into groups until the classification process reaches a terminal node. The decision process can be expressed as

$$f(X) = \begin{cases} C_1, & \text{if condition}_1 \text{ is satisfied,} \\ C_2, & \text{if condition}_2 \text{ is satisfied,} \\ \vdots & \\ C_n, & \text{if condition}_n \text{ is satisfied,} \end{cases} \quad (2)$$

where  $X$  represents the input feature vector,  $f(X)$  represents the decision function, and  $C_1, C_2, \dots, C_n$  represent the possible output classes. The model assigns each observation to a class according to the sequence of decision rules from the root node to a leaf node.

### 3.6.3 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a supervised learning method that classifies an observation based on the classes of its nearest training samples. The algorithm calculates the distance between a test sample and the available training samples using a selected distance metric. The  $k$  nearest samples are then identified, and the class with the highest frequency among these neighbors is assigned to the test sample. Euclidean and Manhattan distances can be used to measure the similarity between observations.

The Euclidean distance between a training sample  $x_i$  and a test sample  $x_u$  can be expressed as

$$d(x_i, x_u) = \sqrt{\sum_{j=1}^p (x_{ij} - x_{uj})^2}, \quad (3)$$

where  $d(x_i, x_u)$  represents the Euclidean distance between the training and test samples,  $p$  represents the number of features,  $x_{ij}$  represents the value of the  $j$ -th feature in the  $i$ -th training sample, and  $x_{uj}$  represents the corresponding feature value in the test sample. The KNN model determines the predicted class from the majority class among the  $k$  nearest training samples.

### 3.6.4 Support Vector Classification

Support Vector Classification (SVC) is a supervised learning method that separates classes using an optimal decision boundary. The model identifies a hyperplane that maximizes the margin between the classes. The observations located closest to the decision boundary are defined as support vectors. These observations determine the position of the separating hyperplane. The decision function of a linear SVC can be expressed as

$$f(X) = \text{sign}(w^T X + b), \quad (4)$$

where  $X$  represents the input feature vector,  $w$  represents the weight vector, and  $b$  represents the bias term. The sign of the decision function determines the predicted class of each observation.

For non-linear classification problems, SVC can transform the input space through a kernel function. This transformation allows the model to construct a decision boundary in a higher-dimensional feature space. In this study, SVC is used as one of the classification models for evaluating stunting malnutrition data.

### 3.6.5 Gradient Boosting Decision Trees

Gradient Boosting Decision Trees (GBDT) is an ensemble learning method that combines multiple decision trees in a sequential process. Each new tree is trained to reduce the prediction error produced by the current ensemble. The final prediction is obtained by combining the outputs of all trees according to their respective contributions. The GBDT prediction can be expressed as

$$F(x) = \sum_{i=1}^N \alpha_i f_i(x), \quad (5)$$

where  $F(x)$  represents the final prediction,  $f_i(x)$  represents the prediction produced by the  $i$ -th decision tree,  $\alpha_i$  represents the contribution of the  $i$ -th tree, and  $N$  represents the total number of trees. The sequential learning process allows each additional tree to update the ensemble based on the prediction errors identified in the previous stage.

### 3.6.6 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is an ensemble learning method that extends gradient boosting through regularization and optimized tree construction. The model builds decision trees sequentially, where each new tree is added to reduce the prediction error produced by the current ensemble. Regularization is applied to control model complexity during the learning process. XGBoost can also provide feature-importance information based on the contribution of input variables to the constructed trees. In this study, XGBoost is used as one of the classification models for evaluating stunting malnutrition data.

### 3.6.7 Light Gradient Boosting Machine

Light Gradient Boosting Machine (LightGBM) is a gradient-boosting method that constructs decision trees using a leaf-wise growth strategy. The algorithm expands the leaf that produces the largest reduction in the objective function during each iteration. LightGBM also supports parallel computation and efficient memory usage during model training. The model can provide feature-importance information based on the contribution of input variables to the constructed trees. In this study, LightGBM is used as one of the classification models for evaluating stunting malnutrition data.

### 3.6.8 Adaptive Boosting

Adaptive Boosting (AdaBoost) is an ensemble learning method that combines several weak learners into a single classification model. During each iteration, the algorithm increases the weights of misclassified samples and decreases the weights of correctly classified samples. This weighting mechanism directs the subsequent learner to focus on observations that were classified incorrectly in the previous iteration. The final prediction is obtained by combining the outputs of all weak learners according to their respective weights. In this study, AdaBoost is used as one of the classification models for evaluating stunting malnutrition data.

### 3.6.9 Categorical Boosting

Categorical Boosting (CatBoost) is a boosting-based machine learning method designed to process categorical features directly during model training. The algorithm reduces the need for explicit one-hot encoding by applying internal transformations to categorical variables. CatBoost constructs the model sequentially, where each new learner is added to reduce the prediction error produced by the previous ensemble. The method also provides information about feature contributions that can be used to examine the role of input variables in the classification process. In this study, CatBoost is used as one of the classification models for evaluating stunting malnutrition data.

### 3.6.10 Stacking Classifier

Stacking Classifier is an ensemble learning method that combines the outputs of multiple base classifiers within a two-level learning structure. In this study, Gradient Boosting Decision Trees, Random Forest, XGBoost, and CatBoost are used as the base classifiers. Each base classifier generates prediction probabilities from the input data. These probabilities are then used as input features for the meta-model. The meta-model combines the

information produced by the base classifiers to generate the final class prediction. In this study, the Stacking Classifier is used as one of the classification models for evaluating stunting malnutrition data.

### 3.7. Grey Wolf Optimization

Grey Wolf Optimization (GWO) is a population-based metaheuristic algorithm inspired by the social hierarchy and hunting behavior of grey wolves. The algorithm represents candidate solutions through four hierarchical levels: alpha ( $\alpha$ ), beta ( $\beta$ ), delta ( $\delta$ ), and omega ( $\omega$ ). The alpha represents the best candidate solution, while the beta and delta represent the second and third best candidate solutions. The remaining search agents are represented by omega and update their positions according to the three leading solutions [25].

The hunting process in GWO consists of searching, encircling, and approaching the prey. The encircling process is modeled using the position of a search agent and the estimated position of the prey [26]. The position update is expressed as

$$\mathbf{X}(t+1) = \mathbf{X}_p(t) - \mathbf{A} \cdot \mathbf{D}, \quad (8)$$

where

$$\mathbf{D} = |\mathbf{C} \cdot \mathbf{X}_p(t) - \mathbf{X}(t)|, \quad (7)$$

$\mathbf{X}(t)$  represents the current position of a search agent,  $\mathbf{X}_p(t)$  represents the estimated position of the prey, and  $t$  represents the current iteration. The coefficient vectors  $\mathbf{A}$  and  $\mathbf{C}$  are calculated as

$$\mathbf{A} = 2a\mathbf{r}_1 - a, \quad (8)$$

and

$$\mathbf{C} = 2\mathbf{r}_2, \quad (9)$$

where  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are random vectors in the interval  $[0, 1]$ , while  $a$  decreases linearly during the optimization process.

The alpha, beta, and delta solutions guide the movement of the remaining search agents. The distances between a search agent and the three leading solutions are calculated as

$$\mathbf{D}_\alpha = |\mathbf{C}_1\mathbf{X}_\alpha - \mathbf{X}|, \quad (10)$$

$$\mathbf{D}_\beta = |\mathbf{C}_2\mathbf{X}_\beta - \mathbf{X}|, \quad (11)$$

and

$$\mathbf{D}_\delta = |\mathbf{C}_3\mathbf{X}_\delta - \mathbf{X}|. \quad (12)$$

The candidate positions generated from the three leading wolves are calculated as

$$\mathbf{X}_1 = \mathbf{X}_\alpha - \mathbf{A}_1\mathbf{D}_\alpha, \quad (13)$$

$$\mathbf{X}_2 = \mathbf{X}_\beta - \mathbf{A}_2\mathbf{D}_\beta, \quad (14)$$

and

$$\mathbf{X}_3 = \mathbf{X}_\delta - \mathbf{A}_3\mathbf{D}_\delta. \quad (15)$$

The new position of a search agent is obtained from the average of the three candidate positions:

$$\mathbf{X}(t+1) = \frac{\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3}{3}. \quad (16)$$

The parameter  $a$  decreases linearly from 2 to 0 during the optimization process. The parameter is calculated as

$$a = 2 - 2\left(\frac{t}{T}\right), \quad (17)$$

where  $t$  represents the current iteration and  $T$  represents the maximum number of iterations. The decreasing value of  $a$  controls the transition between exploration and exploitation during the search process.

### 3.8. Proposed Model

The proposed model integrates GWO with ten machine learning classifiers for feature selection and stunting malnutrition classification. GWO searches the feature space to identify subsets of attributes that support the classification process. Each candidate solution represents a selected feature subset, while the fitness function evaluates the quality of the subset based on classification accuracy and the number of selected features.

The selected feature subsets are evaluated using ten machine learning classifiers: K-Nearest Neighbors, LightGBM, Decision Tree, Support Vector Classifier, Random Forest, XGBoost, AdaBoost, CatBoost, Gradient Boosting Decision Trees, and Stacking Classifier. The Stacking Classifier combines Decision Tree, Random Forest, Gradient Boosting Decision Trees, and XGBoost as base classifiers. The same experimental framework is applied to compare the classifiers under the evaluated feature-selection conditions. The fitness function used to evaluate each candidate feature subset is expressed as

$$\text{Fitness}(t) = - \left[ \alpha \rho + \beta \left( \frac{N - L}{L} \right) \right], \quad (18)$$

where  $\rho$  represents the classification accuracy obtained from the evaluated machine learning model,  $L$  represents the number of selected features, and  $N$  represents the total number of candidate features in the stunting dataset. The parameter  $\alpha$  represents the weight assigned to classification accuracy, while  $\beta$  represents the weight assigned to feature reduction. The parameters satisfy

$$\alpha \in [0, 1], \quad \beta = 1 - \alpha. \quad (19)$$

The optimization process minimizes the fitness value in Equation 18. A candidate solution with a lower fitness value represents a feature subset that provides a better balance between classification accuracy and feature reduction according to the defined objective function. Algorithm 1 presents the GWO-based feature-selection procedure used in the proposed model.

---

**Algorithm 1:** GWO-based feature selection for stunting classification

---

**Input:** Population of search agents  $\mathbf{X}_i, i = 1, 2, \dots, n$

**Output:** Best feature subset  $\mathbf{X}_\alpha$

- 1 Initialize the population  $\mathbf{X}_i$ ;
  - 2 Initialize  $a$ ,  $\mathbf{A}$ , and  $\mathbf{C}$ ;
  - 3 Calculate the fitness of each search agent using Equation 18;
  - 4 Determine  $\mathbf{X}_\alpha$ ,  $\mathbf{X}_\beta$ , and  $\mathbf{X}_\delta$ ;
  - 5 Set  $t \leftarrow 0$ ;
  - 6 **while**  $t < T$  **do**
  - 7     **for**  $i \leftarrow 1$  **to**  $n$  **do**
  - 8         Update the position of  $\mathbf{X}_i$  using Equation 16;
  - 9         Update  $a$ ,  $\mathbf{A}$ , and  $\mathbf{C}$ ;
  - 10        Calculate the fitness of each search agent;
  - 11        Update  $\mathbf{X}_\alpha$ ,  $\mathbf{X}_\beta$ , and  $\mathbf{X}_\delta$ ;
  - 12        Set  $t \leftarrow t + 1$ ;
  - 13 **return**  $\mathbf{X}_\alpha$ ;
- 

### 3.9. Evaluation Models

The performance of the proposed classification models was evaluated using a confusion matrix. The confusion matrix contains four components: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These components were used to calculate accuracy, recall, precision, and F1-score. Accuracy measures the proportion of correctly classified samples among all evaluated samples and is calculated as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (20)$$

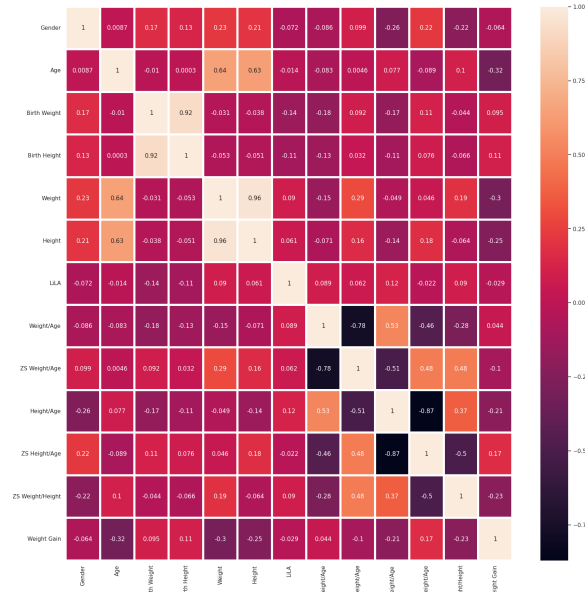


Figure 3. Correlation matrix of the stunting dataset.

Recall measures the proportion of positive samples that are correctly identified and is calculated as

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (21)$$

Precision measures the proportion of correctly predicted positive samples among all samples predicted as positive and is calculated as

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (22)$$

F1-score represents the harmonic mean of precision and recall and is calculated as

$$\text{F1-score} = 2 \left( \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right). \quad (23)$$

These four metrics were used to compare the classification performance of the evaluated models under the different experimental conditions.

## 4. Result and Discussion

### 4.1. Data Distribution and Correlation Analysis

Figure 3 presents the correlation matrix of the variables in the stunting dataset. The matrix shows the direction and strength of the relationships among the analyzed variables. The strongest positive correlation was observed between Weight and Height. This relationship indicates that both variables change in the same direction within the analyzed dataset. Figure 4 presents the distribution of the dataset features according to class membership. The horizontal axis represents the value of each feature, while the vertical axis represents the corresponding density. The gender distribution consists of 51.75% male and 48.25% female observations. The dataset contains two outcome classes, namely Normal and Malnutrition. Before SMOTE was applied, the Normal class represented 54.39% of the observations, while the remaining observations belonged to the Malnutrition class. This distribution shows that the two classes were not equally represented in the original dataset. Figure 5 compares the class distribution before and after SMOTE. After SMOTE was applied, the number of observations in the minority class was increased until the two classes had the same representation. The subsequent experiments compared classification performance before and after class balancing.

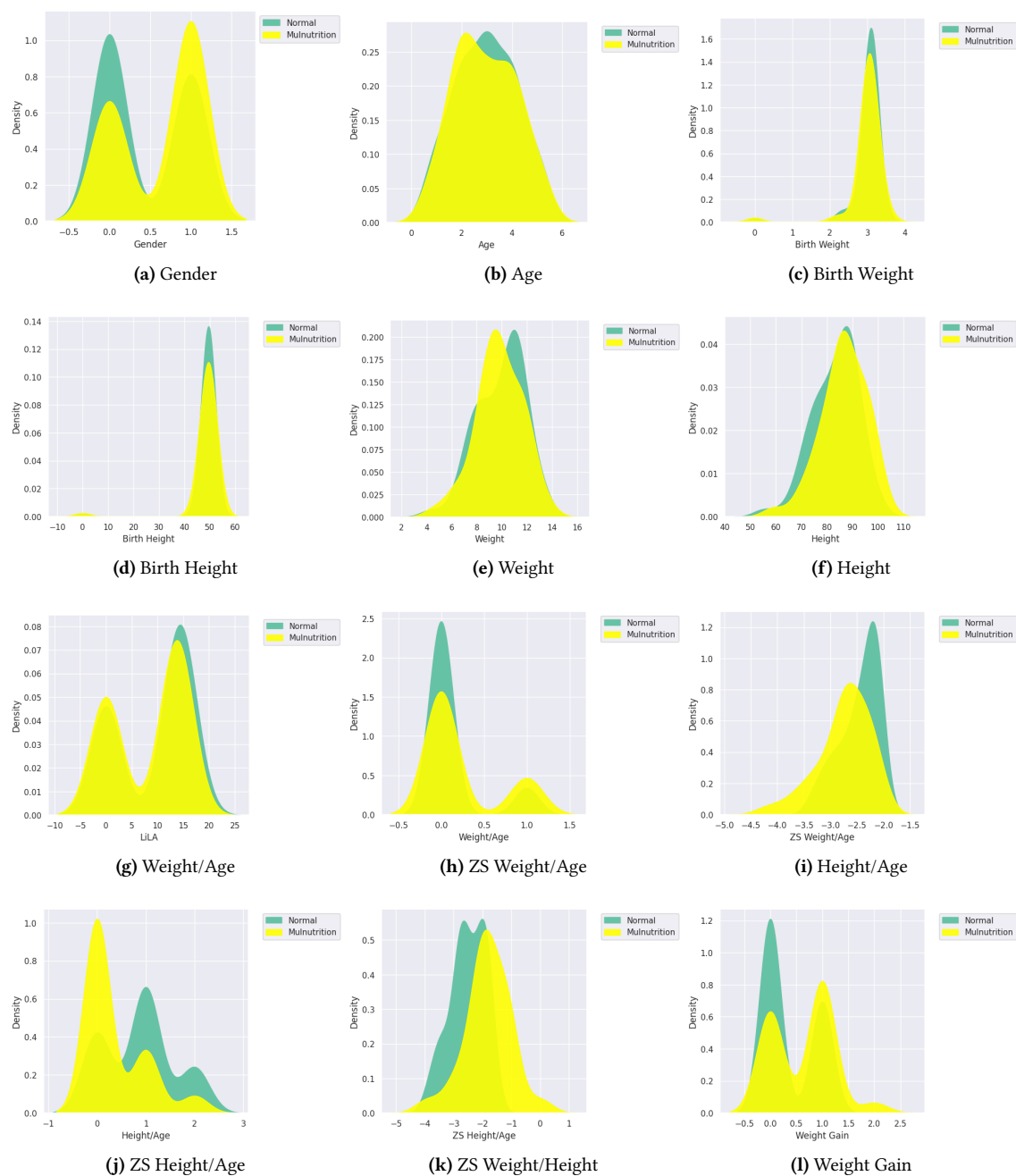


**Figure 4.** Distribution of dataset features according to class membership

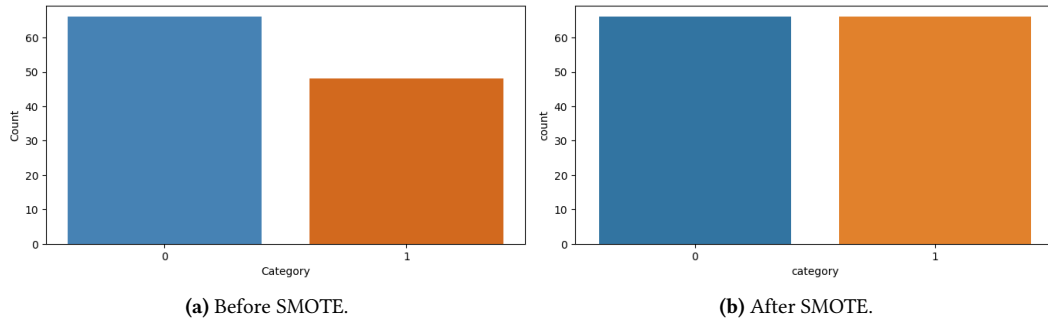


**Figure 5.** Distribution of the classes.

We conducted experiments using 10 different machine learning models on the stunting malnutrition dataset before and after the application of SMOTE to address class imbalance. The model performance results are presented in Figure 6. Figure 6(a) shows that 7 of the 10 models, namely RF, XGBoost, KNN, GBDT, AdaBoost, SVC, and CatBoost, achieved higher accuracy after SMOTE was applied. Accuracy measures the proportion of correctly classified observations. In contrast, DT, LightGBM, and Stacking achieved lower accuracy after SMOTE. These results show that the effect of SMOTE on accuracy differed across the evaluated models.

Figure 6(b) shows that 7 models achieved higher recall after SMOTE, while 3 models achieved lower recall. Recall measures the proportion of positive cases that are correctly identified. The increase in recall indicates that some models identified a larger proportion of positive cases after class balancing. The decrease in recall observed in the remaining models shows that the effect of SMOTE on positive-case identification was model dependent.

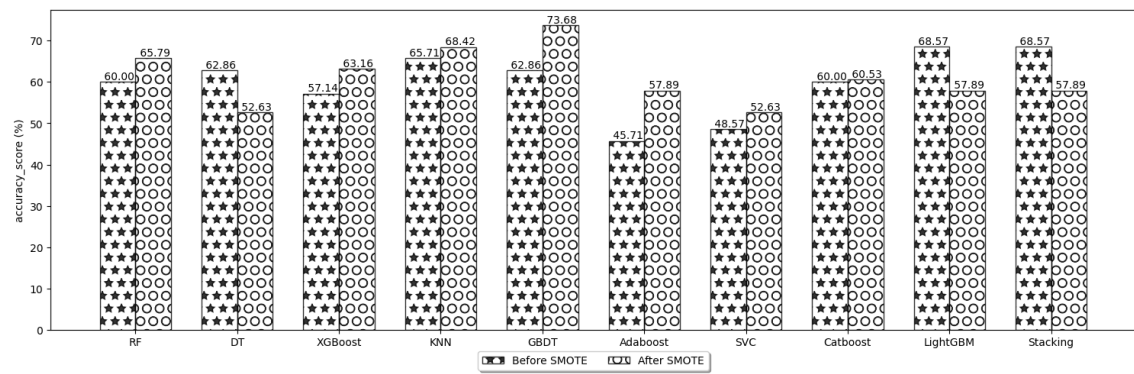
Figure 6(c) and Figure 6(d) show that all models achieved higher precision and F1-score after SMOTE was applied. Precision measures the proportion of correctly predicted positive cases among all positive predictions, while F1-score represents the harmonic mean of precision and recall. These results show that SMOTE improved precision and F1-score across all evaluated models, although its effect on accuracy and recall varied among the classifiers.

The classification performance obtained using all stunting malnutrition attributes indicates that feature selection is required to reduce irrelevant or less informative variables. To address this issue, this study integrates GWO with ten machine learning models for feature selection and classification. GWO searches the feature space to identify attributes that contribute to the classification process. The selected attributes are then used as inputs for the machine learning models. This integration allows the classification process to focus on a reduced set of informative attributes and evaluates whether feature selection improves model performance. The GWO parameter settings used in the feature-selection process are presented in Table 2. The configuration includes the maximum number of iterations, population size, search domain, number of dimensions, and the weights used in the fitness function.

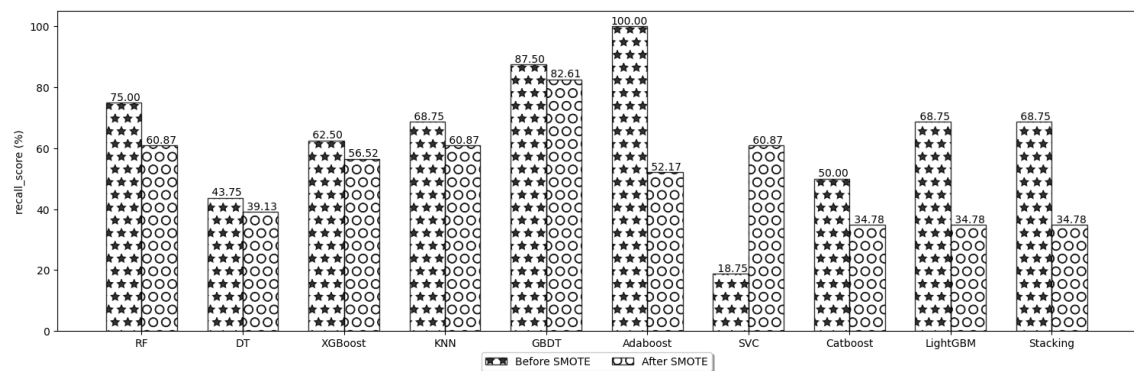
**Table 2.** Parameter settings of the proposed method.

| Parameter                            | Value    |
|--------------------------------------|----------|
| Maximum number of iterations ( $T$ ) | 5        |
| Number of wolves ( $n$ )             | 120      |
| Search domain                        | $[0, 1]$ |
| Number of dimensions ( $N$ )         | 13       |
| Fitness weight ( $\alpha$ )          | 0.4999   |
| Fitness weight ( $\beta$ )           | 0.4999   |

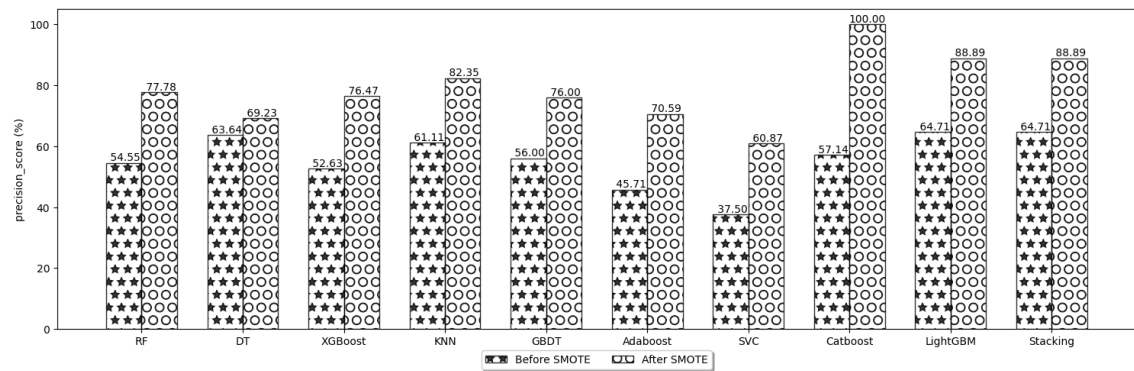
The selected features were obtained by applying GWO to the complete set of stunting malnutrition attributes using ten machine learning models. The parameter settings include the maximum number of iterations, number of wolves, number of dimensions, search domain, and the  $\alpha$  and  $\beta$  weights used in the fitness function, as presented in Table 2. The GWO-based feature-selection process follows Algorithm 1, while the quality of each candidate feature subset is evaluated using the fitness function defined in Equation 18.



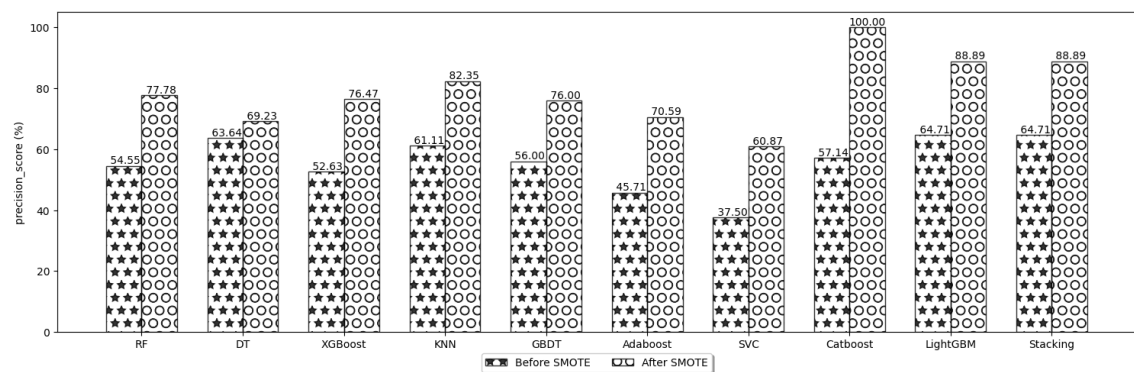
(a) Accuracy.



(b) Recall.



(c) Precision.



(d) F1-score.

Figure 6. Model evaluation with all stunting malnutrition attributes.

The optimization process minimizes the fitness value by updating the positions of the search agents during each iteration. A lower fitness value represents a candidate feature subset with a better objective-function value according to the defined fitness formulation. Figure 7 presents the fitness values obtained by the GWO-based models before and after SMOTE. The decreasing fitness values across the iterations show the progression of the search process toward feature subsets with lower objective-function values.

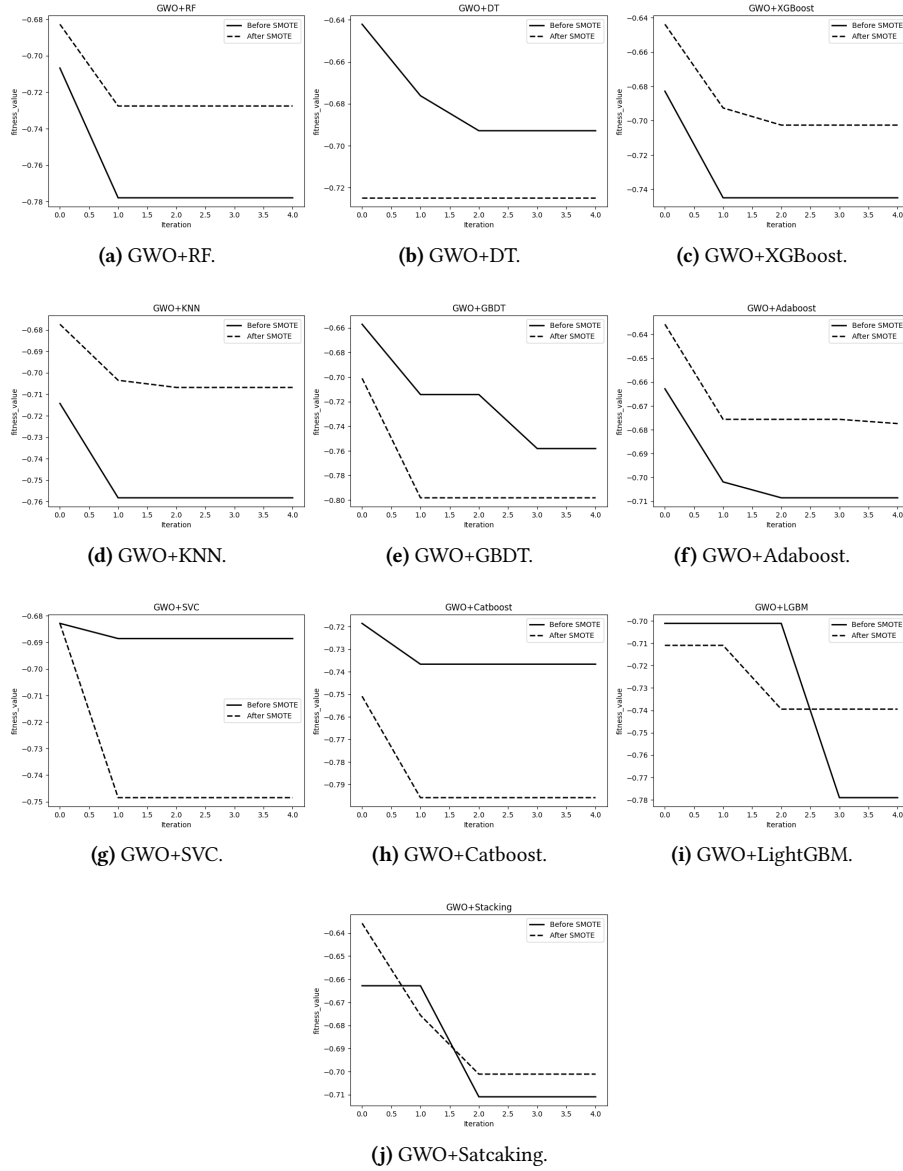
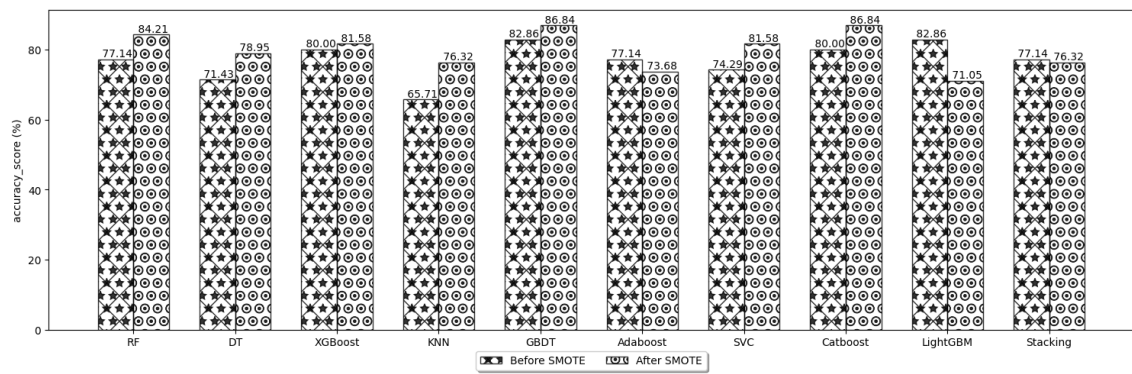


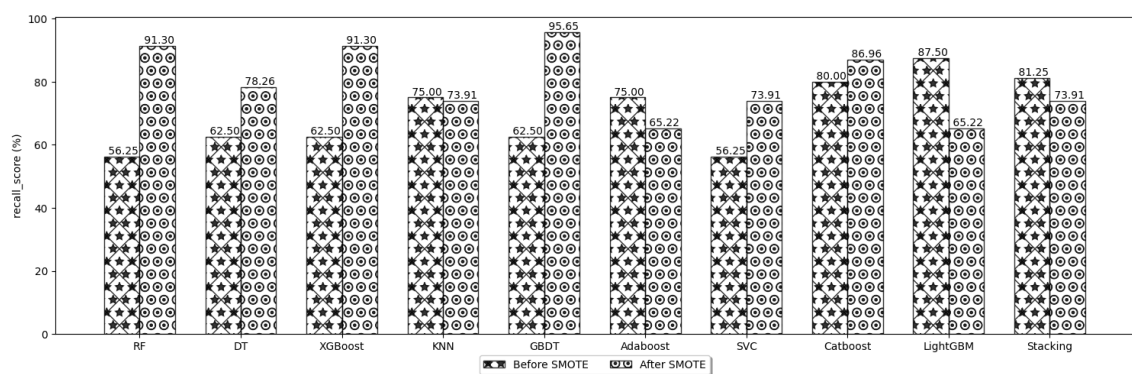
**Figure 7.** The fitness value of the proposed model.

The feature-selection results obtained using GWO with ten machine learning models, before and after SMOTE, are presented in Table 3. Several attributes, including Age, Height, LiLA, W/A, and W/H, were selected repeatedly across multiple model configurations. This pattern indicates that these attributes were frequently retained by the GWO-based feature-selection process under the evaluated experimental conditions. The selected feature subsets differed across classifiers and between the conditions before and after SMOTE. These differences show that the feature-selection results depended on both the classification model and the class-balancing condition. Therefore, the GWO-based selection process identified different combinations of attributes for different model configurations rather than producing a single identical feature subset for all classifiers.

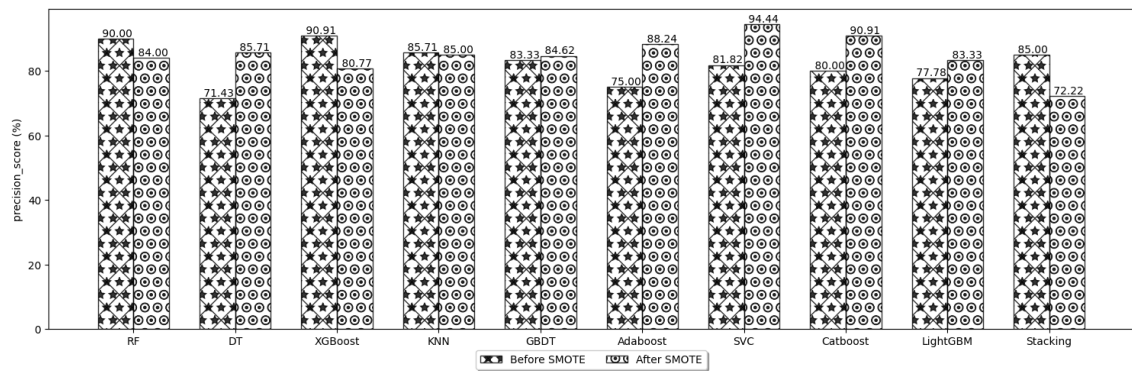
The feature subsets selected by the GWO-based procedure were used as inputs for the ten machine



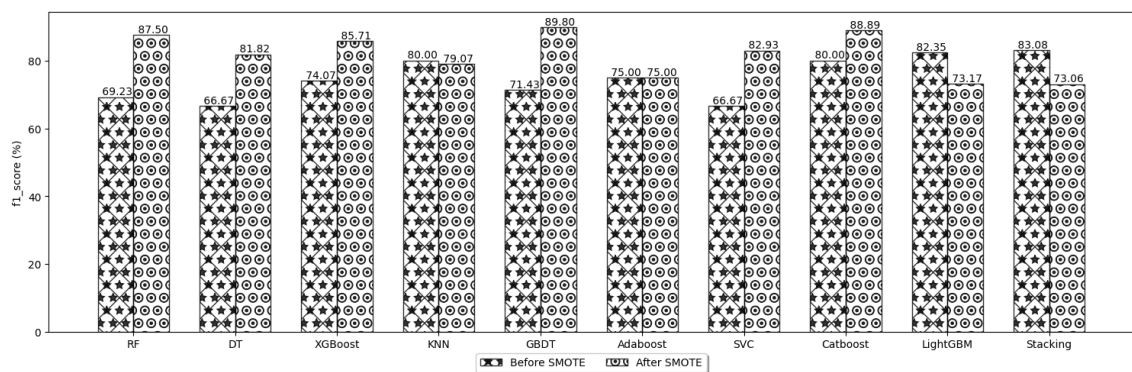
(a) Accuracy.



(b) Recall.



(c) Precision.



(d) F1-score.

**Figure 8.** Model evaluation with feature selection of stunting malnutrition attributes.

**Table 3.** Feature subsets selected by GWO across ten machine learning models before and after SMOTE.

| Model    | Condition | Gender | Age | Birth Height | Birth Weight | Weight | Height | LiLA | W/A | ZS W/A | H/A | ZS H/A | W/H | ZS W/H |
|----------|-----------|--------|-----|--------------|--------------|--------|--------|------|-----|--------|-----|--------|-----|--------|
| RF       | Before    | ✓      | ✓   | ✗            | ✗            | ✗      | ✗      | ✗    | ✗   | ✗      | ✓   | ✗      | ✗   | ✗      |
|          | After     | ✗      | ✓   | ✗            | ✗            | ✗      | ✓      | ✗    | ✗   | ✓      | ✗   | ✗      | ✓   | ✗      |
| DT       | Before    | ✗      | ✓   | ✗            | ✗            | ✗      | ✗      | ✓    | ✗   | ✗      | ✓   | ✗      | ✗   | ✓      |
|          | After     | ✗      | ✓   | ✓            | ✗            | ✗      | ✗      | ✓    | ✓   | ✗      | ✓   | ✓      | ✓   | ✗      |
| XGBoost  | Before    | ✓      | ✓   | ✗            | ✗            | ✗      | ✗      | ✓    | ✓   | ✗      | ✓   | ✗      | ✗   | ✗      |
|          | After     | ✗      | ✓   | ✗            | ✗            | ✗      | ✓      | ✓    | ✗   | ✓      | ✗   | ✗      | ✗   | ✗      |
| KNN      | Before    | ✓      | ✓   | ✓            | ✗            | ✗      | ✗      | ✗    | ✓   | ✓      | ✓   | ✓      | ✓   | ✗      |
|          | After     | ✓      | ✓   | ✗            | ✗            | ✗      | ✗      | ✓    | ✗   | ✓      | ✗   | ✓      | ✗   | ✗      |
| GBDT     | Before    | ✗      | ✓   | ✗            | ✗            | ✗      | ✗      | ✓    | ✗   | ✗      | ✓   | ✗      | ✗   | ✗      |
|          | After     | ✓      | ✗   | ✗            | ✗            | ✓      | ✓      | ✓    | ✗   | ✗      | ✗   | ✗      | ✓   | ✗      |
| AdaBoost | Before    | ✗      | ✓   | ✓            | ✓            | ✗      | ✗      | ✓    | ✓   | ✗      | ✓   | ✗      | ✓   | ✗      |
|          | After     | ✓      | ✓   | ✓            | ✓            | ✓      | ✗      | ✗    | ✗   | ✓      | ✓   | ✗      | ✗   | ✗      |
| SVC      | Before    | ✗      | ✓   | ✗            | ✓            | ✓      | ✗      | ✗    | ✗   | ✓      | ✗   | ✗      | ✗   | ✓      |
|          | After     | ✗      | ✓   | ✓            | ✓            | ✓      | ✗      | ✗    | ✓   | ✓      | ✗   | ✗      | ✗   | ✓      |
| CatBoost | Before    | ✓      | ✓   | ✗            | ✗            | ✗      | ✓      | ✓    | ✗   | ✗      | ✓   | ✗      | ✗   | ✗      |
|          | After     | ✓      | ✓   | ✗            | ✗            | ✗      | ✓      | ✓    | ✗   | ✓      | ✓   | ✗      | ✓   | ✗      |
| LightGBM | Before    | ✗      | ✓   | ✗            | ✗            | ✗      | ✗      | ✗    | ✗   | ✓      | ✓   | ✗      | ✗   | ✗      |
|          | After     | ✗      | ✓   | ✗            | ✗            | ✗      | ✗      | ✗    | ✗   | ✗      | ✗   | ✗      | ✗   | ✗      |
| Stacking | Before    | ✗      | ✓   | ✗            | ✗            | ✗      | ✓      | ✓    | ✓   | ✗      | ✓   | ✗      | ✓   | ✓      |
|          | After     | ✓      | ✗   | ✗            | ✗            | ✓      | ✓      | ✓    | ✓   | ✗      | ✗   | ✗      | ✓   | ✓      |

Note: ✓ = selected feature; ✗ = unselected feature. W/A = weight-for-age; H/A = height-for-age; W/H = weight-for-height; ZS = z-score.

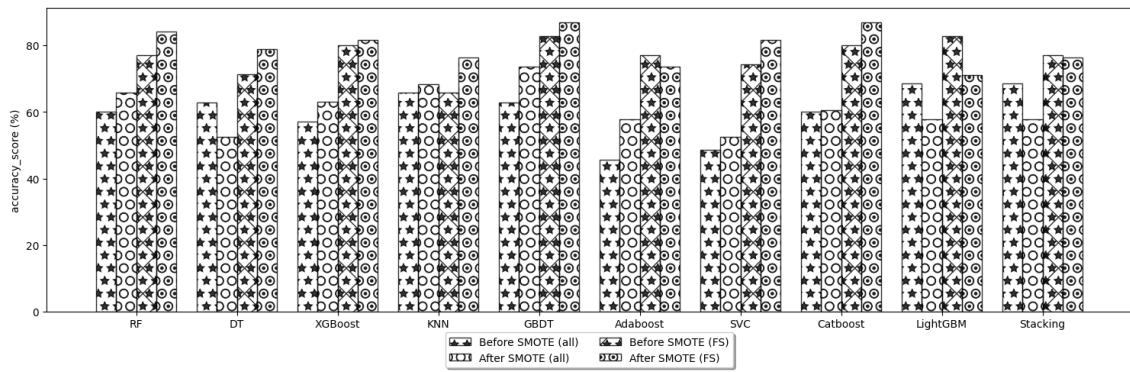
learning models, and the evaluation results are presented in Figure 8. Figure 8(a) shows that 7 of the 10 models achieved higher accuracy after SMOTE was applied. In contrast, Figure 8(b) shows that all models achieved lower recall after SMOTE. Figure 8(c) and Figure 8(d) show that all models achieved higher precision and F1-score after SMOTE.

These results show that the effect of SMOTE differed across evaluation metrics after GWO-based feature selection. The increase in accuracy, precision, and F1-score was accompanied by a decrease in recall across all evaluated models. This pattern indicates a trade-off between overall classification performance and the proportion of positive cases that were correctly identified. Therefore, the effect of SMOTE should be interpreted using multiple evaluation metrics rather than a single performance measure.

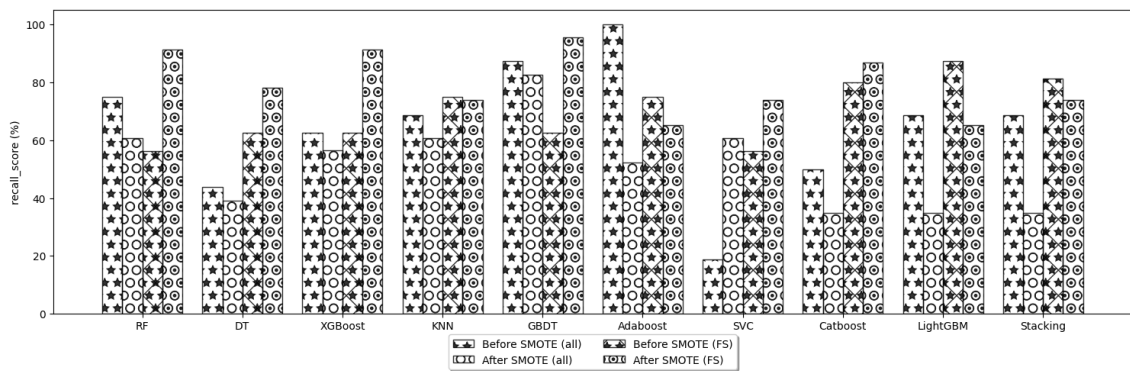
The performance comparison across the experimental conditions is presented in Figure 9. The comparison includes results obtained before and after SMOTE, using both the complete feature set and the GWO-selected feature subsets. Based on accuracy, precision, and F1-score, GBDT achieved the highest values among the evaluated models. In contrast, recall values varied across the models. These results show that model performance depended on the evaluation metric. GBDT produced the highest accuracy, precision, and F1-score under the evaluated conditions, while the recall results did not follow the same pattern. This variation indicates that a model with higher accuracy, precision, and F1-score does not necessarily achieve the highest recall. Therefore, model selection should consider the evaluation metric that is most relevant to the objective of stunting malnutrition classification.

## 5. Conclusion and Future Work

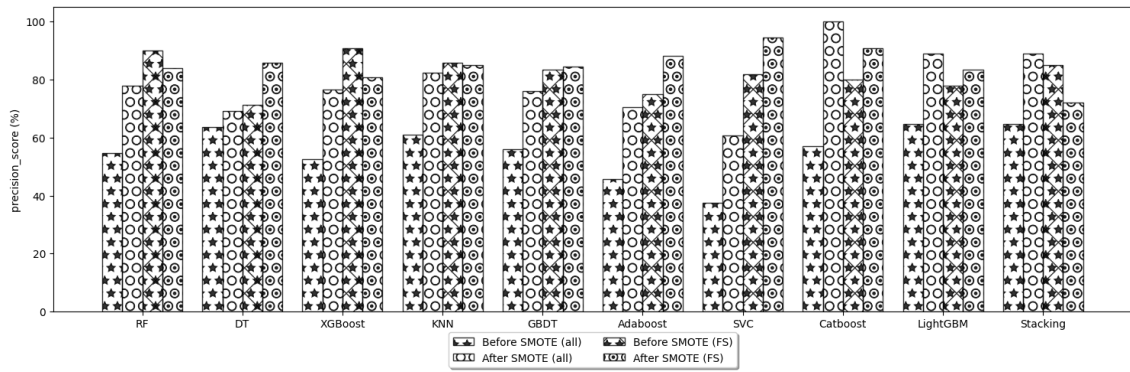
This study evaluated ten machine learning models, namely RF, DT, XGBoost, KNN, GBDT, AdaBoost, SVC, CatBoost, LightGBM, and Stacking, combined with Grey Wolf Optimization (GWO) for stunting malnutrition classification. The experimental results show that the effect of GWO-based feature selection and SMOTE varied across models and evaluation metrics. GBDT achieved the highest accuracy, precision, and F1-score among the evaluated models under the reported experimental conditions. Using all attributes, GBDT achieved an accuracy of 62.86% before SMOTE and 73.68% after SMOTE. After GWO-based feature selection, the accuracy increased to 82.86% before SMOTE and 86.84% after SMOTE. Recall results showed a different pattern across the evaluated models, indicating that higher accuracy, precision, and F1-score did not always correspond to higher recall. These findings show that feature selection and class balancing affected model performance differently across the evaluated metrics. The results also indicate that model assessment should consider multiple performance measures rather than accuracy alone. Future work can extend this study in



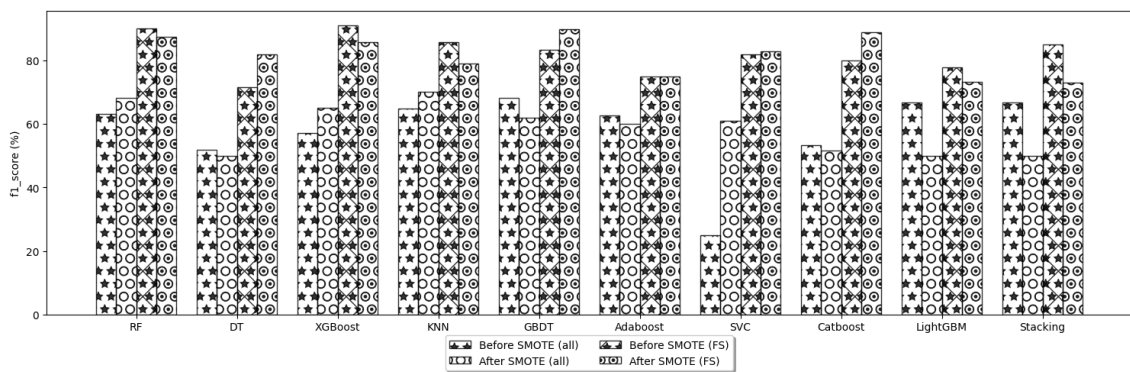
(a) Accuracy.



(b) Recall.



(c) Precision.



(d) F1-score.

Figure 9. The comparison of model evaluation with and without feature selection

several directions. First, additional machine learning models and alternative feature-selection methods can be evaluated using the same dataset. Second, larger and more diverse datasets can be used to examine the consistency of the classification results. Third, external validation can be conducted using stunting datasets from different healthcare facilities or geographical areas. Fourth, hyperparameter optimization can be applied systematically to the evaluated classifiers and the GWO configuration. Fifth, additional clinical variables, including child health history and relevant risk factors, can be incorporated into the dataset. These extensions can provide a broader evaluation of machine learning and feature-selection methods for stunting malnutrition classification.

## Author Contributions

Conceptualization: S. and Z.A.; Methodology: S.; Software: S.; Validation: S. and Z.A.; Formal analysis: S.; Investigation: S.; Resources: S. and Z.A.; Data curation: S.; Writing—original draft preparation: S.; Writing—review and editing: S. and Z.A.; Visualization: S.; Supervision: Z.A.; Project administration: S.; Funding acquisition: S. and Z.A.

## Funding

This research received no external funding.

## Data Availability Statement

The data supporting the findings of this study are not publicly available due to privacy and ethical restrictions related to child health data. Access to the dataset is restricted in accordance with institutional and ethical requirements.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgments

ChatGPT, developed by OpenAI, was used to assist with language editing and manuscript formatting. The authors reviewed all generated content and remain fully responsible for the accuracy and integrity of the manuscript.

## References

- [1] Leroy JL, Frongillo EA. Perspective: What Does Stunting Really Mean? A Critical Review of the Evidence. *Advances in Nutrition*. 2019;10(2):196-204.
- [2] Efevbera Y, Bhabha J, Farmer PE, Fink G. Girl Child Marriage as a Risk Factor for Early Childhood Development and Stunting. *Social Science & Medicine*. 2017;185:91-101.
- [3] Perkins JM, Kim R, Krishna A, McGovern M, Aguayo VM, Subramanian SV. Understanding the Association Between Stunting and Child Development in Low- and Middle-Income Countries: Next Steps for Research and Intervention. *Social Science & Medicine*. 2017;193:101-9.
- [4] World Health Organization. Levels and Trends in Child Malnutrition: 2020 Edition; 2020. Joint Child Malnutrition Estimates. Available from: <https://www.unicef.org/reports/joint-child-malnutrition-estimates-levels-and-trends-child-malnutrition-2020>.
- [5] Kementerian Kesehatan Republik Indonesia. Profil Kesehatan Indonesia. Kementerian Kesehatan Republik Indonesia; 2021.
- [6] Dinas Kesehatan Kota Semarang. Dashboard Kesehatan Dinas Kesehatan Kota Semarang; 2023. Available from: <http://119.2.50.170:9095/dashboardNew/index.php>.

- [7] Rifada M, Chamidah N, Ningrum RA, Muniroh L. Stunting Determinants Among Toddlers in Probolinggo District of Indonesia Using Parametric and Nonparametric Ordinal Logistic Regression Models. *Communications in Mathematical Biology and Neuroscience*. 2023.
- [8] Darnila E, Maryana M, Mawardi K, Sinambela M, Pahendra I. Supervised Models to Predict the Stunting in East Aceh. *International Journal of Engineering, Science and Information Technology*. 2022;2(3):33-9.
- [9] Ndagijimana S, Kabano IH, Masabo E, Ntaganda JM. Prediction of Stunting among Under-5 Children in Rwanda Using Machine Learning Techniques. *Journal of Preventive Medicine and Public Health*. 2023;56(1):41-9.
- [10] Rahman J, Ahmed SMF, Abedin NAMM, Ahammed MB, Ali M, Rahman MJ. Investigate the Risk Factors of Stunting, Wasting, and Underweight among Under-Five Bangladeshi Children and Its Prediction Based on Machine Learning Approach. *PLOS ONE*. 2021;16(6):e0253172. Available from: <https://doi.org/10.1371/journal.pone.0253172>.
- [11] Mburu R. Comparison of Elastic Net and Random Forest in Identifying Risk Factors of Stunting in Children Under Five Years of Age in Kenya; 2020.
- [12] Al-Tashi Q, Md Rais H, Abdulkadir SJ, Mirjalili S, Alhussian H. A Review of Grey Wolf Optimizer-Based Feature Selection Methods for Classification. In: *Evolutionary Machine Learning Techniques. Algorithms for Intelligent Systems*. Singapore: Springer; 2020. p. 273-86. Available from: [https://doi.org/10.1007/978-981-32-9990-0\\_13](https://doi.org/10.1007/978-981-32-9990-0_13).
- [13] Al-Tashi Q, Rais H, Jadid S. Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification. *Advances in Intelligent Systems and Computing*. 2019;843:257-66.
- [14] Karamti H, Alharthi R, Al Anizi A, Alhebshi RM, Eshmawi AA, Alsubai S, et al. Improving Prediction of Cervical Cancer Using KNN Imputed SMOTE Features and Multi-Model Ensemble Learning Approach. *Cancers*. 2023;15(17).
- [15] Khan MNA, Yunus RM. A Hybrid Ensemble Approach to Accelerate the Classification Accuracy for Predicting Malnutrition among Under-Five Children in Sub-Saharan African Countries. *Nutrition*. 2023;108.
- [16] Chilyabanyama ON, Chilengi R, Simuyandi M, Chisenga CC, Chirwa M, Hamusonde K, et al. Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia. *Children*. 2022;9(7).
- [17] Wajgi R, Wajgi D. Malnutrition Detection in Infants Using Machine Learning Approach. In: *AIP Conference Proceedings*; 2022. .
- [18] Hermanto A, Gumelar AB, Ongkodjojo E, Khudrati WC, Young A, Ano Djoka MM, et al. Prediction of Nutritional Requirements for Children's Growth and Adolescents Using Machine Learning. In: *2022 International Seminar on Application for Technology of Information and Communication*; 2022. p. 263-7.
- [19] Jain S, Khanam T, Abedi AJ, Khan AA. Efficient Machine Learning for Malnutrition Prediction among Under-Five Children in India. In: *2022 IEEE Delhi Section Conference*; 2022. .
- [20] Khan JR, Tomal JH, Raheem E. Model and Variable Selection Using Machine Learning Methods with Applications to Childhood Stunting in Bangladesh. *Informatics for Health and Social Care*. 2021;46(4):425-42.
- [21] Harrison E, Syed S, Ehsan L, Iqbal NT, Sadiq K, Umrani F, et al. Machine Learning Model Demonstrates Stunting at Birth and Systemic Inflammatory Biomarkers as Predictors of Subsequent Infant Growth: A Four-Year Prospective Study. *BMC Pediatrics*. 2020;20(1).
- [22] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 2022;16:321-57.
- [23] Majzoobi MM, Namdar S, Najafi-Vosough R, Hajilooi AA, Mahjub H. Prediction of Hepatitis Disease Using Ensemble Learning Methods. *Journal of Preventive Medicine and Hygiene*. 2022;63(3):E424-8.

- [24] Goel E, Abhilasha E. Random Forest: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2017;7(1):251-7.
- [25] Badar AQH. Grey Wolf Optimizer. In: *Evolutionary Optimization Algorithms*. CRC Press; 2021. p. 165-90. Available from: <https://doi.org/10.1201/9781003206477-8>.
- [26] Xue B, Zhang M, Browne WN, Yao X. A Survey on Evolutionary Computation Approaches to Feature Selection. *IEEE Transactions on Evolutionary Computation*. 2016;20(4):606-26.