Hybrid Ensemble Learning with SMOTEENN and Soft Voting for Stunting Risk Prediction: A SHAP-Based Explainable Approach

Nuwairy El Furqany[1,*], Muhammad Subianto[2], Asep Rusyana[3]

[1]*Departement of Informatics, Faculty of Mathematics and Natural Sciences, Syiah Kuala University, Banda Aceh, Indonesia*

[2,3]*Departement of Statistics, Faculty of Mathematics and Natural Sciences, Syiah Kuala University, Banda Aceh, Indonesia*

(Received: April 15, 2025; Revised: July 10, 2025; Accepted: October 14, 2025; Available online: October 22, 2025)

Abstract

Stunting remains a critical public health concern in Indonesia, with long-term consequences for physical growth, cognitive development, and human capital. This study introduces a hybrid machine learning framework to predict household-level stunting risk by integrating Synthetic Minority Over-sampling Technique with Edited Nearest Neighbors (SMOTEENN), soft voting ensemble, and SHapley Additive exPlanations (SHAP). The objective is to enhance both predictive accuracy and interpretability in identifying high-risk households. A dataset of 115,579 household records from West Sumatra, comprising 20 demographic, socioeconomic, health, and housing predictors, was utilized. Preprocessing steps included handling missing values, categorical encoding, and applying SMOTEENN exclusively on the training set to mitigate class imbalance. The baseline models demonstrated limited sensitivity, with XGBoost performing best at 74.56% accuracy and 71.08% F1-score on imbalanced data. After applying SMOTEENN, performance improved substantially, with XGBoost achieving 91.82% accuracy and 91.74% F1-score. Further improvements were obtained through hybridization, where the Random Forest and XGBoost soft voting ensemble reached 91.95% accuracy and 92.46% F1-score, representing a notable gain over individual classifiers. SHAP analysis added interpretability by identifying family members, education level, diverse food consumption, occupation, and drinking water source as dominant predictors of stunting risk. The novelty of this study lies in the integration of SMOTEENN with ensemble learning and SHAP, providing not only robust performance but also transparency in feature contributions. The findings demonstrate that the proposed framework improves sensitivity to minority classes, delivers superior predictive accuracy compared to baseline models, and offers interpretable insights to guide targeted interventions. By combining methodological rigor with explainability, this research contributes a practical decision-support tool for policymakers, supporting early detection of at-risk households and accelerating stunting reduction efforts in Indonesia.

Keywords: Stunting Prediction, Hybrid Machine Learning, SMOTEENN, Soft Voting Ensemble, SHAP Interpretability, Class Imbalance, Public Health, Stunting

1. Introduction

Stunting remains a critical global health challenge with long-term implications for human capital development. In Indonesia, it continues to be a priority issue, with the prevalence in West Sumatra reaching 23.6% in 2024, still above the national reduction target set by the Ministry of Health [1]. Beyond impairing physical growth, stunting also affects cognitive development, ultimately diminishing human resource quality and violating children's rights to optimal growth and development [2]. These pressing concerns highlight the importance of developing accurate and reliable predictive models that can guide more effective public health policies and targeted interventions. However, it must also be emphasized that prediction alone does not directly reduce stunting. The true impact of predictive modeling can only be realized when integrated into broader intervention frameworks, such as nutritional programs, maternal and child health services, and community-based education initiatives.

Despite the potential of machine learning in public health, predictive modeling for stunting often encounters serious challenges due to class imbalance in the data. Households at risk of stunting typically represent a minority compared to those not at risk, leading conventional algorithms to be biased toward the majority class [3]. This bias can result in deceptively high overall accuracy while failing to adequately detect the minority group most in need of intervention

***Corresponding author: Nuwairy El Furqany (nuwairy@gmail.com)**
DOI: https://doi.org/10.47738/jads.v6i4.829

[4]. Addressing this imbalance is therefore crucial to improving the sensitivity and fairness of predictive models in stunting risk classification.

The Synthetic Minority Oversampling Technique (SMOTE) has been widely applied to rebalance imbalanced datasets by generating synthetic minority samples. However, SMOTE alone tends to produce overlapping instances near decision boundaries, which may increase misclassification [5]. To mitigate this drawback, SMOTE can be combined with Edited Nearest Neighbors (ENN), forming the hybrid method SMOTEENN. ENN removes noisy or ambiguous majority-class samples located near the minority boundary, thereby improving class separability and model performance [6]. Prior studies [7] on monkeypox case classification have shown that SMOTEENN significantly improves accuracy and F1-score compared to SMOTE alone, highlighting its suitability for imbalanced health-related datasets such as stunting.

Beyond class rebalancing, further performance improvements can be achieved through ensemble learning techniques. Ensemble methods integrate predictions from multiple base learners, reducing variance and improving robustness compared to individual models [8]. Among these, soft voting ensembles have demonstrated superior performance. For example, prior studies [9] reported accuracy improvements ranging from 3% to 9% compared to individual classifiers, underscoring their ability to capture complementary strengths of diverse models while enhancing detection of minority classes.

Nevertheless, achieving high accuracy is insufficient if models remain opaque to policymakers and health practitioners. In public health contexts, interpretability is essential to ensure that model outputs can be trusted and effectively translated into targeted interventions. SHAP addresses this need by quantifying the contribution of each feature to individual predictions using cooperative game theory principles [10]. By enabling both global and local interpretability, SHAP increases the transparency and practical value of predictive models, allowing them to function not merely as analytical tools but as actionable decision-support systems within stunting reduction programs.

To address these challenges, this study introduces a hybrid machine learning framework that combines SMOTEENN for managing class imbalance, Soft Voting Ensemble (SVE) for robust classification, and SHAP for interpretability. This integrated approach seeks to overcome the limitations of previous studies that treated resampling, ensemble learning, and interpretability in isolation. Specifically, the objectives are twofold: (i) to enhance predictive accuracy and sensitivity in detecting at-risk households, thereby improving the fairness of classification under imbalanced conditions, and (ii) to provide transparent explanations of feature contributions, ensuring that the model's insights can be meaningfully applied in public health decision-making. By bridging technical performance with interpretability, the proposed framework aspires not only to improve risk prediction but also to serve as a practical decision-support tool that can inform targeted interventions, strengthen policy design, and ultimately contribute to accelerating stunting reduction efforts in Indonesia.

2. Literature Review

Machine learning has been increasingly applied in public health research to predict and understand complex phenomena such as childhood stunting. Various algorithms and preprocessing techniques have been explored, ranging from traditional classifiers like Logistic Regression (LR) to more advanced ensemble methods such as Random Forest (RF), Gradient Boosting (GB), and XGBoost [11]. These methods have been tested on different demographic and health survey datasets across countries, highlighting both their potential and their limitations in handling class imbalance and ensuring model interpretability.

While the reported findings are promising, prior research has often focused on either feature engineering or resampling in isolation. For instance, studies utilizing feature selection improved accuracy but did not fully address class imbalance [12], while those applying resampling techniques such as SMOTE enhanced balance but sometimes suffered from overlapping synthetic samples near decision boundaries [13], [14]. Similarly, ensemble methods such as RF and GB showed strong predictive performance, yet their application was rarely combined with advanced resampling to further improve sensitivity in minority class detection. Another limitation is the minimal emphasis on interpretability, which is critical in public health applications where model transparency is essential to translate predictions into actionable

interventions. To provide a clearer overview of prior efforts in this domain, table 1 presents a summary of recent studies on stunting risk prediction using machine learning.

Table 1. Previous Studies On Machine Learning For Stunting Risk Prediction

Researcher	Method	Dataset	Preprocessing	Main Findings
[11]	LR, RF, SVM, NB, XGBoost, GB	Rwanda Demographic and Health Survey (RDHS), 4052 samples	Feature selection	Best model GB achieved 80.49% accuracy
[12]	LR, CTree, XGBoost, SVM-RBF	Papua New Guinea Demographic and Health Survey (PNG DHS), 19.200 samples	Feature selection (LASSO, RF-RFE)	Best model LASSO-XGBoost achieved 72.8% accuracy
[13]	XGBoost, RF, SVM, KNN	Kaggle stunting dataset (Indonesia), 10.000 samples	SMOTE	XGBoost with SMOTE achieved 85.74% accuracy
[14]	SVM, GNB, LR, DT, RF, LGB, XGB, KNN	Ethiopian Demographic and Health Survey (EDHS), 15.683 samples	SMOTE + feature selection	RF with SMOTE and feature selection achieved 77% accuracy
This study	LR, RF, SVM, XGBoost, SVE	Stunting risk dataset, West Sumatra (Indonesia), 115.579 samples	SMOTEENN	Improved accuracy on imbalanced data and enhanced interpretability

As shown in table 1, prior research has provided valuable insights into the application of machine learning for stunting risk prediction across multiple contexts. However, most studies emphasized either feature engineering or resampling in isolation, with limited integration of both strategies, and only minimal attention to interpretability. To overcome these gaps, the present study proposes a hybrid framework that combines SMOTEENN to effectively manage class imbalance, SVE to enhance classification robustness, and SHAP for model interpretability, thereby offering a more comprehensive and transparent approach for stunting risk prediction.

3. Methodology

To provide a clear overview of the research process, the workflow of this study is illustrated in figure 1. This workflow outlines the sequential steps taken, starting from dataset preparation to model interpretation, ensuring a systematic approach toward stunting risk prediction.

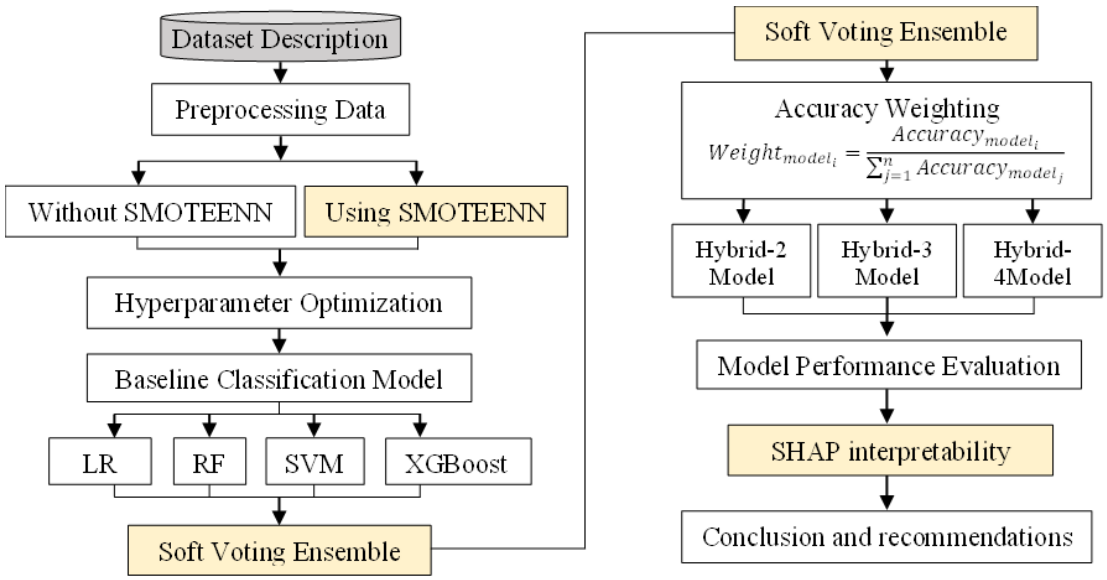

Figure 1. Research Workflow

Figure 1 illustrates the overall workflow of this research, which begins with dataset description and preprocessing to ensure data readiness. The next step involves the development of baseline classification models both without and with SMOTEENN, where SMOTEENN is applied to address the issue of class imbalance. Each classifier, including Logistic

 ISSN 2723-6471

Regression, Random Forest, Support Vector Machine, and XGBoost, undergoes hyperparameter optimization to enhance predictive performance. The optimized models are then integrated through a soft voting ensemble scheme, in which accuracy-based weighting is used to construct hybrid models with different classifier combinations. Following model construction, performance evaluation is conducted using multiple metrics to assess classification effectiveness. Finally, SHAP analysis is employed to interpret the model and identify the most influential features contributing to stunting risk prediction.

3.1. Dataset Description

This study employed secondary data derived from the *Pemutakhiran Pendataan Keluarga* (family data updating survey) conducted by the National Population and Family Planning Board (BKKBN), West Sumatra Province, in 2023. The details of each variable are presented in table 2.

Table 2. Dataset Description

No	Feature	Type	Description
1	Family Members (X_1)	Ratio	Number of family members
2	Marital Status (X_2)	Nominal	1 = Single; 2 = Married; 3 = Divorced; 4 = Widowed
3	Mother's Age at First Marriage (X_3)	Ratio	Age of mother at first marriage
4	Occupation (X_4)	Nominal	1 = Unemployed; 2 = Farmer; 3 = Fisherman; 4 = Trader; 5 = Civil Servant; 6 = Private Employee; 7 = Daily Laborer
5	Education Level (X_5)	Ordinal	1 = No Education; 2 = Elementary; 3 = Junior High; 4 = Senior High; 5 = Higher Education
6	Health Insurance Type (X_6)	Nominal	1 = BPJS-PBI; 2 = BPJS-Non PBI; 3 = Private; 4 = None
7	Regular Worship (X_7)	Nominal	1 = Yes; 0 = No
8	Income Source Ownership (X_8)	Nominal	1 = Yes; 0 = No
9	Diverse Food Consumption (X_9)	Nominal	1 = Yes; 0 = No
10	Asset Ownership (X_{10})	Nominal	1 = Yes; 0 = No
11	Participation in Social Activities (X_{11})	Nominal	1 = Yes; 0 = No
12	Main Lighting Source (X_{12})	Nominal	1 = Yes; 0 = No
13	Cooking Fuel (X_{13})	Nominal	1 = Yes; 0 = No
14	Family Planning Education (X_{14})	Nominal	1 = Yes; 0 = No
15	Online Media Access (X_{15})	Nominal	1 = Yes; 0 = No
16	Floor Condition (X_{16})	Nominal	1 = Decent; 0 = Not Decent
17	Roof Condition (X_{17})	Nominal	1 = Decent; 0 = Not Decent
18	Drinking Water Source (X_{18})	Nominal	1 = Decent; 0 = Not Decent
19	House Condition (X_{19})	Nominal	1 = Decent; 0 = Not Decent
20	Toilet Sanitation (X_{20})	Nominal	1 = Decent; 0 = Not Decent
21	Stunting Risk (Y)	Nominal	1 = At Risk; 0 = Not at Risk

The dataset provides comprehensive information on family-level social, economic, demographic characteristics, and stunting risk indicators. The initial population comprised 132,921 households, of which 115,579 valid records were retained after data cleaning. Each record corresponds to one household and includes multiple socio-demographic and economic attributes. Binary variables (coded as 1 = Yes and 0 = No) were obtained from structured household interviews conducted by trained enumerators, with responses validated through BKKBN's quality control procedures. This large-scale dataset offers a unique advantage by representing diverse socio-economic and demographic conditions across West Sumatra, making it highly suitable for developing and validating predictive models in public health research [15].

3.2. Data Preprocessing

3.2.1. Handling Missing Values

Before modeling was conducted, the data underwent a preprocessing stage to ensure quality, consistency, and readiness for training. This step is crucial to prevent bias and prediction errors that may arise from inconsistent scales or formats within the dataset [16]. The first step involved handling missing values, where rows or columns with a large proportion of missing entries and deemed unrepresentative for analysis were removed to maintain data quality. The choice of missing value treatment method was based on the data distribution, the proportion of missing entries, and their potential impact on the analysis, thereby minimizing bias in the model's results.

3.2.2. Feature Encoding

In this study, label encoding was applied to categorical attributes to convert text-based responses into numerical representations so that they could be processed by machine learning algorithms. For nominal variables, the encoding served purely as identifiers without implying any order or ranking among categories [17]. For instance, binary attributes such as Online Media Access or Asset Ownership were encoded into 1 = Yes and 0 = No, while multi-category variables such as Occupation or Health Insurance Type were assigned integer codes to represent distinct categories. For ordinal variables such as Education Level, the encoded values reflected the inherent order (e.g., 1 = No Education to 5 = Higher Education). This approach ensured that the encoding captured both the nominal meaning of unordered variables and the ordinal meaning where applicable, without introducing artificial relationships.

3.2.3. SMOTEENN for Handling Imbalanced Data

At this stage, the preprocessed data were subjected to class balancing techniques to address the issue of imbalanced distribution between stunting risk and non-risk households. This study adopted the SMOTEENN method, which combines the oversampling strategy of the SMOTE with the undersampling approach of ENN. SMOTE works by generating synthetic minority class samples based on feature-space similarities between existing minority instances, thereby increasing their representation within the dataset [18]. However, synthetic oversampling alone may lead to class overlapping near decision boundaries. To mitigate this, ENN is applied subsequently to remove noisy or misclassified majority-class samples, refining the dataset and enhancing class separability. The hybridization of SMOTE and ENN ensures a more balanced, cleaner dataset that improves model sensitivity toward minority cases while maintaining generalization performance [19]. The sequential process of SMOTEENN applied in this study is summarized in table 3.

Table 3. The SMOTEENN Algorithm Applied In This Study

Phase	Process
Input	Dataset x, Consists of majority samples (x_{maj}) and minority samples (x_{min}).
SMOTE (oversampling)	1. Determine oversampling ratio (IR). 2. For each minority sample x_{min}: - Compute Euclidean distance to other minority samples. - Identify k_1 nearest neighbors. - For $I = 1, \dots, IR$, randomly select one neighbor and generate a new synthetic instance x_{new} by interpolation between x_{min} and its neighbor. 3. Add x_{new} to the minority class.
ENN (undersampling)	For each sample x_i in the SMOTE dataset (minority and majority): - Compute Euclidean distance to other samples. - Identify k_2 nearest neighbors. - Determine the majority label among neighbors. - If the label of x_i differs from the majority, remove x_i (considered as noisy). - Otherwise, retain x_i.
Output	Resampled dataset x', Balanced through SMOTE oversampling and cleaned from outliers or noisy samples using ENN.

In this study, the oversampling ratio (IR) and the number of nearest neighbors (k) were determined through a combination of empirical testing and validation. The IR was adaptively set to achieve a target balance of 1:1 between minority and majority classes within each training split, ensuring that the minority class was sufficiently represented before applying ENN. For the SMOTE phase, the number of nearest neighbors was tuned within the range $k_1 \in \{3,5,7\}$ using inner 5-fold cross-validation, with $k_1 = 5$ yielding the best trade-off between recall and F1-score for the minority class. For the ENN phase, we compared $k_2 \in \{3,5\}$ and selected $k_2 = 3$, as this setting effectively removed noisy samples near decision boundaries while preserving sensitivity to minority cases. These hyperparameters were applied exclusively on the training data, with the test set kept untouched to avoid data leakage.

3.3. Machine Learning Algorithms

3.3.1. Logistic Regression

LR is a widely used statistical method for binary classification. It maps input values to a probability ranging from 0 to 1, estimating the likelihood that a given instance belongs to the positive class (1) rather than the negative class (0). The model is constructed by fitting a linear equation with the input features, but instead of predicting the binary outcome directly, it models the natural logarithm of the odds of the positive class. This value is then transformed into a probability through the logistic, or Sigmoid function, producing an output between 0 and 1 that represents the predicted likelihood [20]. Mathematically, the LR method can be expressed as follows.

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}} \quad (1)$$

$P(Y = 1|X)$ represents the probability of the dependent variable Y being equal to 1 given the input features X, and $(\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ are the model parameters that need to be estimated from the training data.

3.3.2. Random Forest

Random Forest is one of the methods belonging to the decision tree family. This approach constructs multiple decision trees consisting of root nodes, internal nodes, and leaf nodes, where both attribute selection and data sampling are performed randomly according to specific rules. The root node functions as the starting point of data processing, the internal nodes represent questions or splitting criteria based on attributes, while the leaf nodes indicate the final outcome of the decision or classification [21]. The process begins by randomly selecting a training subset from the overall training dataset. Each decision tree in the forest is generated and trained using this subset.

$$\hat{y} = \frac{1}{N_{trees}} \sum_{i=1}^{N_{trees}} y_i \quad (2)$$

$\hat{y}$ represents the final prediction or output of the random forest model. The variable N_{trees} denotes the total number of decision trees within the forest, and each y_i corresponds to the individual prediction made by the i -th decision tree.

3.3.3. Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm designed to classify data by mapping it into a high-dimensional feature space using a nonlinear mapping function [22]. The training data, represented as vectors $\vec{x}_i$, is classified into two categories, denoted as $\vec{y}_i$, which can take the values -1 or 1, as shown in the formula:

$$G = (\vec{x}_i \in \mathbb{R}^n; \vec{y}_i = -1 \; or \; 1; i = 1,2, \dots, N) \quad (3)$$

The goal of SVM is to find the optimal hyperplane that best separates the two classes in this feature space. The algorithm begins by identifying the points in each class that are closest to the separating hyperplane these points are known as the support vectors [22]. Once the support vectors are determined, the distance between the hyperplane and these points is calculated. This distance is called the margin, and the primary objective of SVM is to maximize the margin. A larger margin indicates better generalization and separation between the classes, thus producing a more robust classifier. By maximizing the margin, SVM effectively minimizes the classification error on both the training and unseen data, ensuring high classification performance.

3.3.4. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is an algorithm developed from the Gradient Boosting Decision Tree (GBDT) framework and is designed to efficiently build boosted trees with parallel execution capability. As an ensemble learning technique, XGBoost has proven highly effective for solving both classification and regression problems. The algorithm constructs an ensemble of decision trees sequentially, where each new tree attempts to correct the residual errors of the previous ones using the gradient boosting approach. In the regression tree structure employed by XGBoost, the internal nodes represent attribute test values, while the leaf nodes contain scores that reflect the final decision [23]. The final prediction is obtained by summing the scores generated by all K trees, as expressed in the following equation (4).

$$\text{Loss}_{XGBoost} = \sum_{n=1}^{N} L(\hat{y}_n, y_n) + \sum_{m=1}^{M} \Omega(f_m), \tag{4}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2. \tag{5}$$

XGBoost is a machine learning algorithm that aims to minimize a loss function, $\text{Loss}_{XGBoost}$, which quantifies the error between model predictions and actual values. It operates on a dataset represented as a sequence of data points, X_n, where each data point includes a feature vector and a target value. The discrepancy between model predictions and actual values for each data point is measured using a specific loss function, $L(\hat{y}_n, y_n)$. To control model complexity and prevent overfitting, XGBoost employs the term $\Omega(f)$, consisting of two components: γT, which is related to the number of leaves in the model, and $\lambda \|w\|^2$, where λ is a control constant and w is a vector.

3.4. Soft Voting Ensemble for Hybrid Model

Voting ensemble is one of the most common forms of ensemble learning, a machine learning approach that combines multiple models to improve prediction accuracy and stability. The main objective of this method is to produce a final model that outperforms the performance of each individual classifier [24]. By aggregating the strengths of different algorithms, ensemble learning can capture both linear and non-linear relationships in the data, making it a powerful aggregation technique that has been widely adopted in statistics and machine learning for more than a decade [25].

Unlike hard voting, which relies solely on majority rule, soft voting leverages probabilistic weighting to better capture the confidence levels of each model, thereby reducing variance and improving robustness [26]. This hybrid approach allows the system to exploit the complementary strengths of linear and non-linear classifiers, resulting in improved accuracy and sensitivity, particularly in detecting minority cases of stunting risk. In soft voting, the final class label of the response variable is determined based on the predicted probability p from each base classifier. The ensemble prediction is obtained using the following equation (6).

$$\hat{y} = \arg\max \sum_{j=1}^{M} w_j p_j(c) \tag{6}$$

where $\hat{y}$ denotes the predicted class label, M is the total number of classifiers, $p_j(c)$ is the probability predicted by the j-th classifier for class c, and w_j is the weight assigned to the j-th classifier, which can be adjusted according to its performance. The novelty of this study lies in the weighting mechanism of the base classifiers in the soft voting ensemble. Unlike conventional soft voting where model weights are set arbitrarily or equally, in this research the weights were assigned proportionally according to the accuracy of each model relative to the combined accuracy of all models [26]. The weighting scheme is defined as follows equation (7).

$$Weight_{model_i} = \frac{Accuracy_{model_i}}{\sum_{j=1}^{n} Accuracy_{model_j}} \tag{7}$$

In this study, accuracy was chosen as the weighting criterion for the soft voting ensemble because it provides a straightforward and consistent measure across all base models, thereby facilitating direct comparison and integration. Models that produce more reliable predictions are assigned higher weights to ensure that the final result is not

disproportionately influenced by any single model [26]. The combinations consist of baseline individual models, hybrid models with two, three, and four classifiers, as summarized in table 4.

Table 4. Hybrid Machine Learning Model Combinations

No	Model Type	Number of Models	Hybrid Model Combinations
1	Baseline	4	LR, RF, SVM, XGBoost
2	Hybrid-2	6	LR+RF, LR+SVM, LR+XGBoost, RF+SVM, RF+XGBoost, SVM+XGBoost
3	Hybrid-3	4	LR+RF+SVM, LR+RF+XGBoost, LR+SVM+XGBoost, RF+SVM+XGBoost
4	Hybrid-4	1	LR+RF+SVM+XGBoost

The hybrid model combinations in table 4 were selected based on diversity and performance considerations rather than random choice. LR, RF, SVM, and XGBoost were chosen as baseline models because they represent different learning paradigms. Hybrid ensembles were then constructed by systematically combining these models to leverage their complementary strengths, in line with the principle that diversity among classifiers improves ensemble performance [26].

3.5. Evaluation Metrics

The actual data and the predicted results from the classification model are represented using a cross-tabulation known as the confusion matrix. This matrix provides detailed information about the relationship between the true class labels (rows) and the predicted class labels (columns), enabling evaluation of the model's classification performance [27]. The structure of the confusion matrix is shown in table 5.

Table 5. Confusion Matrix

Real Class	Predicted Positive Class	Predicted Negative Class
Real Positive Class	True Positive (TP)	False Negative (FN)
Real Negative Class	False Positive (FP)	True Negative (TN)

As shown in table 5, the confusion matrix summarizes model performance by cross-tabulating actual versus predicted classes: true positives (TP) are households correctly identified as at risk of stunting, false positives (FP) are non-risk households incorrectly labeled as at risk, true negatives (TN) are households correctly identified as not at risk, and false negatives (FN) are at-risk households missed by the model. Based on this matrix, precision expresses how reliable positive predictions are (the share of predicted at-risk households that are truly at risk), while recall or sensitivity captures how completely the model identifies actual at-risk households (the share of truly at-risk households that the model flags). The F1-score provides a single, balanced indicator that harmonizes precision and recall, which is especially useful under class imbalance. Finally, accuracy reflects the overall proportion of correct classifications across both at-risk and not-at-risk households.

3.6. SHAP for Interpretability Model

To ensure that the predictive model is not only accurate but also transparent and interpretable, this study employed SHAP as a post-hoc interpretability technique. SHAP computes the contribution of each feature to the prediction outcome in an additive manner, thereby enabling interpretation at both global and local levels [28]. SHAP is an explainability method for machine learning models that is grounded in cooperative game theory through the use of Shapley values. The main idea of SHAP is to calculate the Shapley value for each feature of a given instance, where each value represents the marginal impact of that feature on the prediction [29]. A model prediction can be represented using shapley as follows:

$$\hat{y}_i = shap_0 + shap(X_{1i}) + shap(X_{2i}) + \cdots + shap(X_{ji}) \quad (8)$$

where $\hat{y}_i$ denotes the prediction for the i-th instance, $shap_0 = \mathbb{E}(\hat{y}_i)$ represents the global average prediction, and $shap(X_{ji})$ indicates the contribution of the j-th feature for the i-th instance. This additive decomposition allows each prediction to be explained as the sum of a global baseline and the marginal contributions of individual features, thereby providing both global and local interpretability of the model [30].

4. Results and Discussion

4.1. Data Preprocessing

Prior to model construction, several preprocessing steps were applied to ensure data quality and consistency. First, records with incomplete or invalid information were removed, resulting in 115,579 valid household entries from the initial population of 132,921. Next, categorical variables were transformed into numerical representations using label encoding, allowing both nominal and ordinal predictors to be processed by machine learning algorithms. After encoding, the dataset was divided into training and testing subsets with an 80:20 stratified split, ensuring that the class distribution of stunting risk was preserved in both sets. To address the class imbalance, the SMOTEENN technique was applied exclusively to the training data, combining oversampling of the minority class with noise reduction from the majority class. This approach ensured that the testing data remained in its original distribution, thereby providing an unbiased evaluation of the model's predictive performance. The application of resampling techniques significantly changed the class distribution in the training dataset, while the test dataset was kept in its original form to ensure unbiased evaluation. Table 6 shows the comparison of class distributions.

Table 6. Number of Observations BEFORE and After SMOTEENN

		Train	Percent	Test	Percent
Original	Risk stunting	32696	35.36%	8174	35.36%
	Non-risk stunting	59767	64.64%	14942	64.64%
	All	92463	100.00%	23116	100.00%
SMOTE	Risk stunting	59767	50.00%	8174	35.36%
	Non-risk stunting	59767	50.00%	14942	64.64%
	All	92463	100.00%	23116	100.00%
SMOTEENN	Risk stunting	32886	47.08%	8174	35.36%
	Non-risk stunting	36959	52.92%	14942	64.64%
	All	69845	100.00%	23116	100.00%

As shown in table 6, SMOTE balances both classes exactly at 50:50, whereas SMOTEENN not only oversamples the minority class but also removes noisy samples from the majority class, resulting in a slightly imbalanced but cleaner dataset (47.08% vs. 52.92%). A chi-square test of independence was performed to assess changes in class distribution before and after resampling. The results indicated a highly significant difference across the three datasets ($\chi^2 = 4780.18$, $p < 0.05$). This finding confirms that both SMOTE and SMOTEENN substantially altered the distribution of stunting risk classes compared to the original dataset. In other words, the rebalancing procedures not only adjusted class proportions but also introduced a statistically significant shift, thereby supporting the validity of the preprocessing step. The changes in class distribution before and after the application of SMOTEENN can be visualized in figure 2.

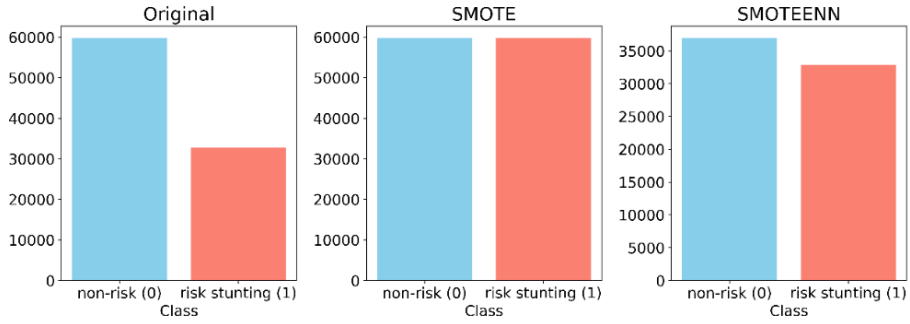

Figure 2. Comparison of Class Distributions: Original, SMOTE, and SMOTEENN

4.2. Performance Without SMOTEENN

Before evaluating the classification models, hyperparameter optimization was carried out using cross-validation to obtain the best configuration for each algorithm. Table 7 summarizes the range of hyperparameters tested, the optimal values selected, and the corresponding cross-validation scores.

Table 7. Best Hyperparameter For Models

Real Class	Hyperparameter	Range	Best Hyperparameter	CV Score
LR	C	[0.01, 0.1, 1, 10, 100]	C = 0.01	0.7213
RF	n_estimator	[100, 200, 500]	n_estimator = 100	0.7388
	max_depth	[5, 10, 20]	max_depth = 5	
	min_samples_split	[2, 5, 10]	min_samples_split = 10	
SVM	C	[0.01, 0.1, 1, 10, 100]	C = 1	0.7261
	kernel	['rbf', 'poly', 'linear']	kernel = 'rbf'	
XGBoost	n_estimator	[100, 200, 500]	n_estimator = 100	0.7428
	max_depth	[3, 5, 10]	max_depth = 5	
	learning rate	[0.01, 0.1, 0.3]	learning rate = 0.1	

Using the optimized configurations, all models were first trained on the original imbalanced dataset without the application of SMOTEENN. The evaluation outcomes, as presented in table 8, reveal that although the overall accuracy values ranged between 72%–75%, this metric alone is misleading given the skewed class distribution. A closer inspection shows that the recall values were consistently lower across all models, highlighting a significant limitation in their ability to correctly identify households at risk of stunting.

Table 8. Performance of Baseline Models without SMOTEENN

Model	Accuracy	Precision	Recall	F1 Score
LR	72.49	70.26	66.50	67.28
RF	74.00	72.09	68.49	69.37
SVM	72.89	72.00	65.46	66.18
XGBoost	74.56	72.27	70.45	71.08

This suggests that the models were biased toward the majority class, producing acceptable accuracy at the expense of failing to capture minority class instances, which are the primary focus of the study. Among the baseline models, XGBoost demonstrated the most promising performance, reaching an accuracy of 74.56% and an F1-score of 71.08, outperforming the other algorithms in balancing precision and recall. Nonetheless, the relatively low recall underscores the necessity of addressing class imbalance through advanced resampling techniques such as SMOTEENN to improve the detection of high-risk households. The details of model misclassifications can be observed in the confusion matrix, as presented in figure 3.

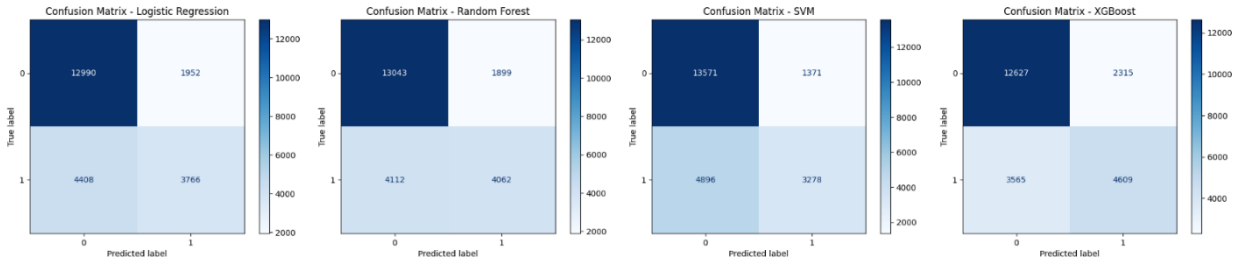

Figure 3. Confusion Matrix without SMOTEENN

4.3. Performance After Using SMOTEENN

In the second experiment, model training was conducted using the SMOTEENN technique applied to the training dataset in order to address the issue of class imbalance. This approach was expected to improve the models' ability to detect minority cases (households at risk of stunting) while maintaining overall predictive performance. After applying SMOTEENN on the training dataset, the performance of the classification models showed a substantial improvement compared to the baseline scenario. As presented in table 9, all models achieved higher accuracy, precision, recall, and F1-scores, indicating that the class balancing procedure significantly enhanced the ability of the models to correctly identify households at risk of stunting.

ISSN 2723-6471

Table 9. Performance of Models with SMOTEENN

Model	Accuracy	Precision	Recall	F1 Score
LR	85.78	85.74	85.73	85.73
RF	91.72	91.73	91.64	91.68
SVM	87.61	87.57	87.56	87.56
XGBoost	91.82	91.84	91.74	91.74

The ensemble-based methods demonstrated the strongest performance, with Random Forest reaching an accuracy of 91.72% and XGBoost slightly outperforming with 91.82%. The consistent improvement across all metrics highlights the effectiveness of SMOTEENN in handling class imbalance, leading to more robust and reliable predictions. These improvements are further illustrated in the confusion matrix after applying SMOTEENN, as shown in figure 4.

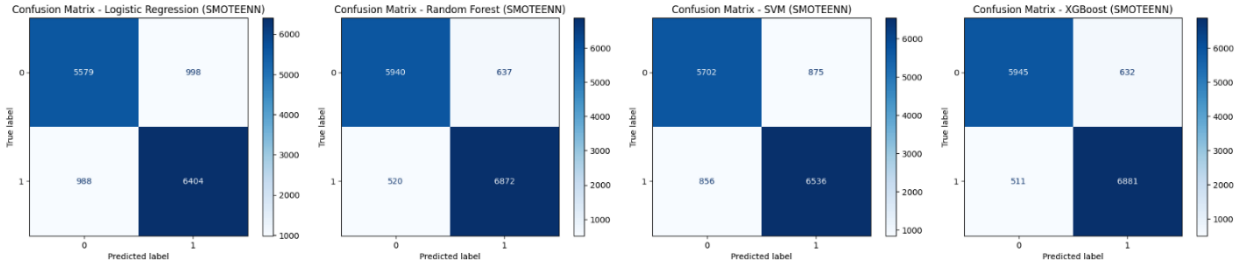

Figure 4. Confusion Matrix with SMOTEENN

4.4. Performance Hybrid Model Soft Voting Ensemble

In the final stage of experimentation, the optimized base classifiers were combined using the SVE approach. Unlike the baseline models, this ensemble method assigned weights to each classifier proportionally to its accuracy, ensuring that models with stronger predictive ability contributed more to the final decision. The evaluation results for individual models, hybrid-2, hybrid-3, and hybrid-4 ensembles are summarized in table 10.

Table 10. Performance of Hybrid Models with Soft Voting Ensemble

No	Models	Weight	Quantity	Accuracy	Precision	Recall	F1 Score
1	LR	[1.00]	1	85.78	85.74	85.73	85.73
2	RF	[1.00]	1	91.72	91.73	91.64	91.68
3	SVM	[1.00]	1	87.61	87.57	87.56	87.56
4	XGBoost	[1.00]	1	91.82	91.84	91.74	91.74
5	LR, RF	[0.48328, 0.51672]	2	89.77	90.10	90.62	90.36
6	LR, SVM	[0.49474, 0.50526]	2	86.87	87.62	87.57	87.59
7	LR, XGBoost	[0.48301, 0.51699]	2	90.68	90.92	91.53	91.22
8	RF, SVM	[0.51146, 0.48854]	2	90.16	90.26	91.25	90.75
9	RF, XGBoost	[0.49973, 0.50027]	2	91.95	91.72	93.21	92.46
10	SVM, XGBoost	[0.48827, 0.51173]	2	90.82	90.92	91.83	91.37
11	LR, RF, SVM	[0.32358, 0.34596, 0.33046]	3	88.88	89.13	89.95	89.54
12	LR, RF, XGBoost	[0.31852, 0.34055, 0.34093]	3	91.32	91.29	92.41	91.85
13	LR, SVM, XGBoost	[0.32345, 0.33034, 0.34621]	3	89.56	89.88	90.46	90.17
14	RF, SVM, XGBoost	[0.33826, 0.32311, 0.33863]	3	91.39	91.41	92.41	91.91
15	LR, RF, SVM, XGBoost	[0.24034, 0.25696, 0.24545, 0.25725]	4	90.48	90.65	91.44	91.04

The performance results of the hybrid models using the soft voting ensemble are shown in table 10. Compared to individual classifiers, the ensemble combinations generally produced higher scores across all metrics, confirming the effectiveness of integrating multiple models. Among the hybrid-2 models, the combination of Random Forest and

 ISSN 2723-6471
3000

XGBoost achieved the highest performance with an accuracy of 91.95% and an F1-score of 92.46, outperforming both models individually. This indicates that the complementary strengths of RF and XGBoost significantly enhanced the classification results. For the hybrid-3 models, the best performance was obtained from the RF, SVM, and XGBoost combination, yielding an accuracy of 91.39% and an F1-score of 91.91, which was slightly higher than the LR-inclusive hybrids. Meanwhile, the full hybrid model combining all four classifiers (LR, RF, SVM, and XGBoost) achieved an accuracy of 90.48% and an F1-score of 91.04, which, although strong, did not surpass the simpler RF+XGBoost hybrid. These findings suggest that while adding more classifiers does not always guarantee superior performance, selecting complementary models with strong individual accuracies such as RF and XGBoost can lead to the most effective hybrid ensemble.

To highlight the impact of data balancing and ensemble techniques, table 11 presents the comparison of the best-performing models across the three experimental stages: baseline, SMOTEENN, and hybrid soft voting ensemble. As shown, the baseline model using XGBoost achieved an accuracy of 74.56% and an F1-score of 71.08, which improved substantially after applying SMOTEENN, reaching 91.82% accuracy and 91.74 F1-score. The highest performance was obtained by the RF+XGBoost hybrid model, which achieved an accuracy of 91.95% and an F1-score of 92.46.

Table 11. Comparison of Best Models Across different Scenarios

Scenario	Best Model	Accuracy	Precision	Recall	F1 Score
Baseline (Original)	XGBoost	74.56	72.27	70.45	71.08
With SMOTEENN	XGBoost	91.82	91.84	91.74	91.74
Hybrid (SVE)	RF + XGBoost	91.95	91.72	93.21	92.46

These results demonstrate that SMOTEENN was highly effective in addressing class imbalance, while the soft voting ensemble further enhanced predictive performance by leveraging the complementary strengths of Random Forest and XGBoost. Overall, the proposed hybrid approach outperformed individual classifiers, providing a robust and reliable framework for stunting risk prediction. To further validate the performance improvements, a paired t-test was conducted across cross-validation folds comparing the baseline model and the hybrid SVE. The results revealed that the hybrid ensemble achieved significantly higher accuracy, precision, recall, and F1-score than the baseline model ($t = -17.32$, $p < 0.05$), confirming the robustness of the observed gains.

4.5. SHAP for Interpretability Model

The SHAP summary plot visualizes the relative importance and impact of features on model predictions can be visualized in figure 5.

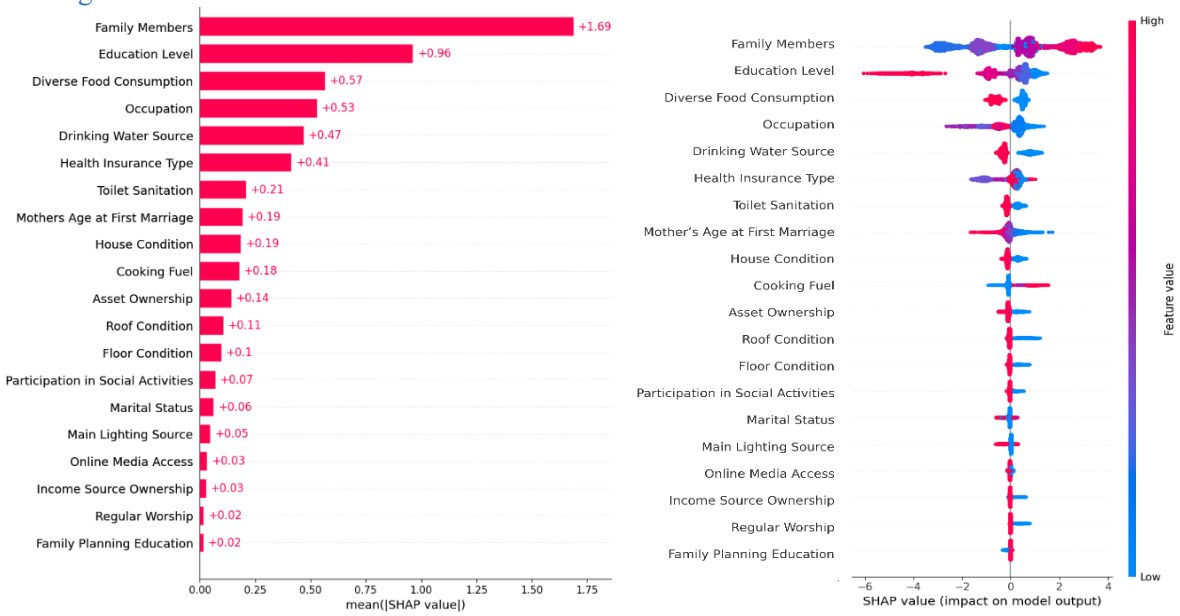

Figure 5. SHAP Summary Plot of Feature Contributions to Stunting Risk Prediction

 ISSN 2723-6471

Figure 5 presents the SHAP summary plot, which combines global and local interpretability of the model. The left panel shows the mean absolute SHAP values for each feature, ranking them by overall importance in predicting stunting risk, while the right panel illustrates the distribution of SHAP values for individual households. The results demonstrate that Family Members (+1.69), Education Level (+0.96), and Diverse Food Consumption (+0.57) are the most influential predictors, followed by Occupation (+0.53) and Drinking Water Source (+0.47). Larger family size and lower maternal education exert strong positive contributions, pushing predictions toward higher stunting risk. Likewise, lack of dietary diversity and high-risk occupations such as farming also increase risk.

The beeswarm distribution in the right panel provides additional nuance by showing how variations in feature values affect the prediction direction. For example, higher values of family size (red) are consistently associated with positive SHAP values, indicating elevated risk, while smaller family sizes (blue) push predictions toward lower risk. Similarly, lower education levels and absence of dietary diversity are clustered on the positive side, reinforcing their role as strong drivers of risk. Conversely, features such as House Condition, Drinking Water Source, and Toilet Sanitation appear on the negative side of the distribution when they are adequate or safe, highlighting their protective influence in reducing stunting risk. Taken together, the summary plot underscores that socio-demographic (family size, education), nutritional (dietary diversity), and environmental (housing and sanitation) factors collectively shape household stunting risk. Beyond validating known determinants, the SHAP results quantify their relative importance and reveal interaction effects, providing a transparent foundation for prioritizing interventions in stunting reduction programs.

While the SHAP summary plot provides a global perspective of the most influential features across the dataset, the SHAP force plot offers local interpretability by explaining predictions for individual households. This visualization decomposes a single prediction into the baseline value and the contributions of each feature, allowing us to see precisely which factors increase or decrease the risk classification. Figure 6 presents SHAP force plots for two households, illustrating how the model decomposes an individual prediction into contributions from each feature.

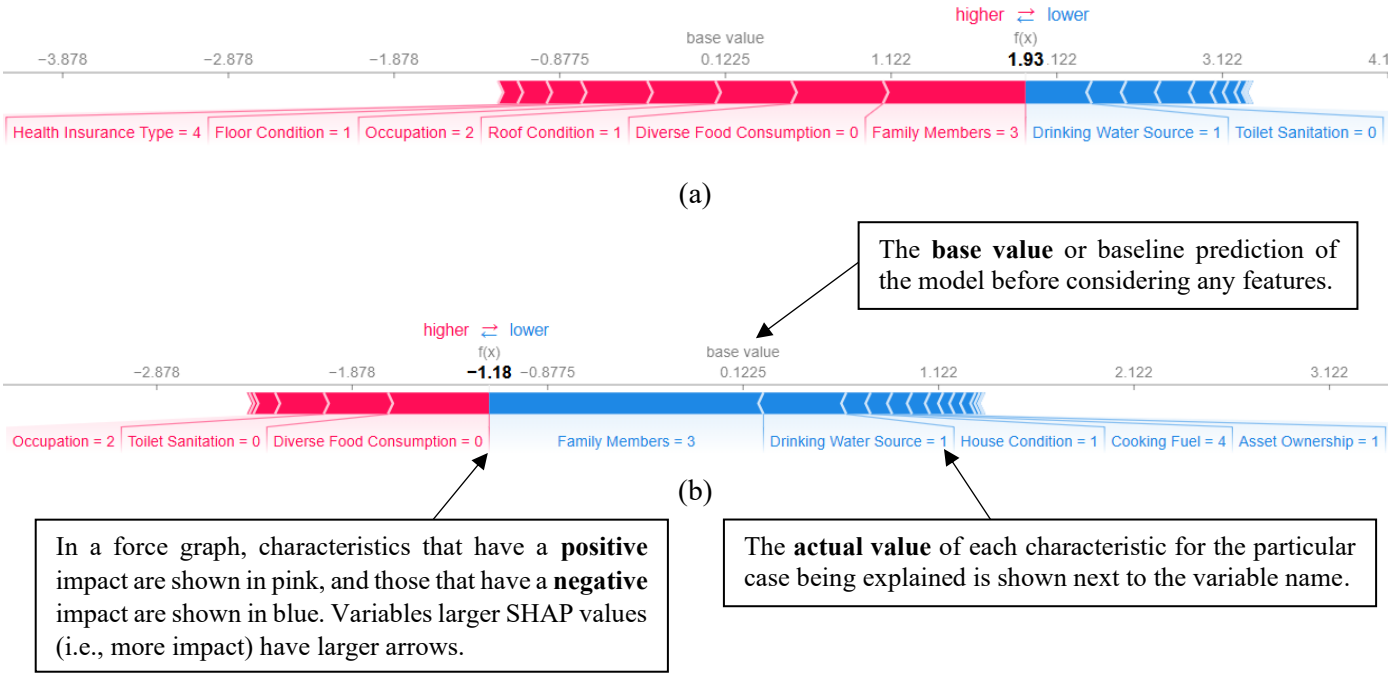

Figure 6. SHAP Force Plot: (A) Sample #10 Classified As At Risk, (B) Sample #100 Classified as Not at Risk

The base value (0.1225) represents the average model output, expressed in log-odds, before accounting for household-specific characteristics. Each arrow indicates the direction and magnitude of a feature's influence, with red arrows pushing the prediction toward higher risk and blue arrows pulling it toward lower risk. The length of the arrow corresponds to the relative strength of the feature's impact.

For sample #10 (figure 6a), the model output reached f(x) = 1.93 (log-odds), which after logistic transformation corresponds to a high predicted probability of being at risk. This upward shift was primarily driven by absence of health

insurance (type 4), farmer occupation, poor dietary diversity, and relatively small family size. Although protective factors such as safe drinking water and adequate toilet sanitation exerted negative contributions, their effect was insufficient to counterbalance the stronger positive drivers, leading to a classification of "at risk."

In contrast, sample #100 (figure 6b) produced a model output of f(x) = –1.18 (log-odds), corresponding to a low probability of stunting risk. Here, protective features such as access to safe drinking water, adequate housing condition, clean cooking fuel, and asset ownership exerted dominant downward effects. While risk-enhancing factors like farmer occupation and poor dietary diversity were present, their contributions were outweighed by stronger protective influences, resulting in the classification of "not at risk."

These household-level explanations underscore SHAP's utility in providing local interpretability. Beyond identifying which households are classified as at risk, the method reveals the precise factors driving each decision. Such insights are valuable for policymakers and practitioners, as they enable the design of targeted interventions—for example, improving dietary diversity or ensuring access to clean water—that directly address the drivers of vulnerability in specific households.

5. Conclusion

This study introduced a hybrid machine learning framework for stunting risk prediction that integrates SMOTEENN for addressing class imbalance, soft voting ensemble for robust classification, and SHAP for model interpretability. The findings confirmed that class imbalance severely constrained baseline model performance, with the best standalone model (XGBoost) achieving only 74.56% accuracy and 71.08% F1-score, underscoring the challenge of detecting minority-class households. By applying SMOTEENN, predictive performance improved substantially across all models, with XGBoost reaching 91.82% accuracy and 91.74% F1-score, demonstrating the critical role of resampling in imbalanced health datasets.

Further enhancement was achieved through the soft voting ensemble strategy, where the hybrid of Random Forest and XGBoost yielded the highest performance (91.95% accuracy and 92.46% F1-score). This result highlights the value of combining complementary learners to improve both accuracy and sensitivity. Beyond predictive strength, SHAP analysis provided interpretability by identifying family size, maternal education, food diversity, occupation, and housing conditions as the most influential predictors, ensuring that the model's decisions remain transparent and actionable.

Taken together, the proposed hybrid framework demonstrates both technical robustness and practical relevance. While the results suggest strong potential for policy applications, actual deployment would require integration with existing health information systems, consideration of data privacy and ethical safeguards, and alignment with community-based intervention strategies. With these considerations addressed, the framework could serve as a decision-support tool to facilitate early detection of at-risk households, guide resource allocation, and support evidence-based programs to reduce stunting prevalence and improve child health outcomes in vulnerable populations.

6. Declarations

6.1. Author Contributions

Conceptualization: N.E.F., M.S., and A.R.; Methodology: N.E.F.; Software: N.E.F.; Validation: M.S., and A.R.; Formal Analysis: M.S., and A.R.; Investigation: M.S.; Resources: A.R.; Data Curation: N.E.F.; Writing Original Draft Preparation: N.E.F.; Writing Review and Editing: M.S., and A.R.; Visualization: N.E.F.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This research was supported by the Kominfo Digital Scholarship (KOMDIGI), a scholarship program organized by the Ministry of Communication and Informatics of the Republic of Indonesia. The researcher would like to express sincere gratitude to the program and all parties who provided valuable support and information throughout this study.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ministry of Health of the Republic of Indonesia, Indonesia Nutritional Status Survey (SSGI) 2024, Jakarta, Indonesia: *Health Development Policy Agency*, 2024.

[2] A. T. Mulyani, M. A. Khairinisa, A. Khatib, and A. Y. Chaerunisaa, "Understanding Stunting: Impact, Causes, and Strategy to Accelerate Stunting Reduction—A Narrative Review," *Nutrients*, vol. 17, no. 9, pp. 1493, 2025. doi: 10.3390/nu17091493.

[3] G. Nduwayezu, P. Zhao, P. Pilesjö, J. P. Bizimana, and A. Mansourian, "Multilevel small-area childhood stunting risk estimation: Insights from spatial ensemble learning, agro-ecological and environmentally remotely sensed indicators," *Environmental and Sustainability Indicators*, vol. 27, no. 1, pp. 100822, 2025. doi: 10.1016/j.indic.2025.100822.

[4] M. Ayele, G. A. Baye, S. H. Yesuf, et al., "Predicting stunting status among under-five children in Ethiopia using ensemble machine learning algorithms," *Scientific Reports*, vol. 15, no. 1, pp. 27907, 2025. doi: 10.1038/s41598-025-03206-1.

[5] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning from Imbalanced Data Sets, Cham, Switzerland: *Springer International Publishing*, vol. 1, no. 1, pp. 377, 2018. doi: 10.1007/978-3-319-98074-4.

[6] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explorations*, vol. 6, no. 1, pp. 20–29, Jun. 2004. doi: 10.1145/1007730.1007735.

[7] L. Siena, T. H. Saragih, R. A. Nugroho, D. Kartini, Muliadi, and W. Caesarendra, "Evaluation of the Impact of SMOTEENN on Monkeypox Case Classification Performance Using Boosting Algorithms," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, pp. 203–220, 2025. doi: 10.35882/ijeeemi.v7i2.77.

[8] T. G. Dietterich, "Ensemble methods in machine learning," in Multiple Classifier Systems, G. Goos, J. Hartmanis, and J. Van Leeuwen, Eds., Springer, Berlin, *Heidelberg*, vol. 1857, no. 1, pp. 1–15, 2000. doi: 10.1007/3-540-45014-9_1.

[9] M. Nguyen, K. Cao-Van, L. G. Minh, T. X. Bui, and S. Hong, "Hybrid Machine Learning Models Using Soft Voting Classifier for Financial Distress Prediction," *SSRN Electronic Journal*, vol. 1, no. 1, pp. 1–19, 2024. doi: 10.2139/ssrn.4941751.

[10] N. Gadde, A. Mohapatra, D. Tallapragada, K. Mody, N. Vijay, and A. Gottumukhala, "Explainable AI for dynamic ensemble models in high-stakes decision-making," *International Journal of Science and Research Archive*, vol. 13, no. 2, pp. 1170–1176, 2024. doi: 10.30574/ijsra.2024.13.2.2091.

[11] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, "Prediction of stunting among under-5 children in Rwanda using machine learning techniques," *Journal of Preventive Medicine & Public Health*, vol. 56, no. 1, pp. 41–49, Jan. 2023. doi: 10.3961/jpmph.22.388.

[12] H. Shen, H. Zhao, and Y. Jiang, "Machine learning algorithms for predicting stunting among under-five children in Papua New Guinea," *Children*, vol. 10, no. 10, pp. 1638, 2023. doi: 10.3390/children10101638.

[13] T. Sugihartono, B. Wijaya, Marini, A. F. Alkayes, and H. A. Anugrah, "Optimizing stunting detection through SMOTE and machine learning: A comparative study of XGBoost, Random Forest, SVM, and k-NN," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 667–682, 2024. doi: 10.47738/jads.v6i1.494.

[14] A. B. Zemariam, B. B. Abate, A. W. Alamaw, E. S. Lake, G. Yilak, M. Ayele, B. D. Tilahun, and H. S. Ngusie, "Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine learning algorithms," *PLoS One*, vol. 20, no. 1, pp. 1–27, 2025. doi: 10.1371/journal.pone.0316452.

[15] National Population and Family Planning Board of the Republic of Indonesia (BKKBN RI), Publication of Families at Risk of Stunting Data 2024, Jakarta, Indonesia: Directorate of Reporting and Statistics, 2024.

[16] A. Tawakuli and T. Engel, "Make your data fair: A survey of data preprocessing techniques that address biases in data towards fair AI," *Journal of Engineering Research*, vol. 1, no. 1, pp. 1–8, 2024. doi: 10.1016/j.jer.2024.06.016.

[17] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y.-I. Cho, "Effective methods of categorical data encoding for artificial intelligence algorithms," *Mathematics*, vol. 12, no. 16, pp. 2553, 2024. doi: 10.3390/math12162553.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 9, pp. 321–357, 2002. doi: 10.1613/jair.953.

[19] G. Husain, D. Nasef, R. Jose, J. Mayer, M. Bekbolatova, T. Devine, and M. Toma, "SMOTE vs. SMOTEENN: A study on the performance of resampling algorithms for addressing class imbalance in regression models," *Algorithms*, vol. 18, no. 1, pp. 37, 2025. doi: 10.3390/a18010037.

[20] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression, 2nd ed., New York, NY, USA: John Wiley & Sons, vol. 2, no. 1, pp. 383, 2000. doi: 10.1002/0471722146.

[21] E. Métais, F. Meziane, M. Saraee, V. Sugumaran, and S. Vadera, Eds., Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems (NLDB 2016), Salford, UK, June 22–24, 2016, Proceedings, Cham, Switzerland: *Springer International Publishing*, vol. 9612, no. 1, pp. 488, 2016. doi: 10.1007/978-3-319-41754-7.

[22] R. G. Brereton and G. R. Lloyd, "Support vector machines for classification and regression," *The Analyst*, vol. 135, no. 2, pp. 230–267, 2010. doi: 10.1039/B918972F.

[23] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD *International Conference on Knowledge Discovery and Data Mining (KDD '16),* San Francisco, CA, USA, vol. 16, no. 1, pp. 785–794, 2016. doi: 10.1145/2939672.2939785.

[24] P. Chithuloori and J. M. Kim, "Soft voting ensemble classifier for liquefaction prediction based on SPT data," *Artificial Intelligence Review*, vol. 58, no. 8, pp. 228, 2025. doi: 10.1007/s10462-025-11230-w.

[25] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: A review," *Healthcare*, vol. 11, no. 12, pp. 1808, 2023. doi: 10.3390/healthcare11121808.

[26] K. Cao-Van, T. Cao Minh, L. Gia Minh, T. B. Q. Thi, and H. Minh Tan, "Soft-voting ensemble model: An efficient learning approach for predictive prostate cancer risk," *Vietnam Journal of Computer Science*, vol. 11, no. 4, pp. 531–552, 2024. doi: 10.1142/S2196888824500155.

[27] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 3rd ed., Burlington, MA, USA: Morgan Kaufmann Publishers, vol. 3, no. 1, pp. 327–391, 2012. doi: 10.1016/C2009-0-61819-5.

[28] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS 30),* Red Hook, NY, USA: Curran Associates, Inc., vol. 2, no. 1, pp. 4765–477, 2017. doi: 10.48550/arXiv.1705.07874.

[29] M. Li, H. Sun, Y. Huang, et al., "Shapley value: from cooperative game to explainable artificial intelligence," *Autonomous Intelligent Systems*, vol. 4, no. 2, pp. 1–12, 2024. doi: 10.1007/s43684-023-00060-8.

[30] A. S. Antonini, J. Tanzola, L. Asiain, G. R. Ferracutti, S. M. Castro, E. A. Bjerg, and M. L. Ganuza, "Machine Learning model interpretability using SHAP values: Application to Igneous Rock Classification task," *Applied Computing and Geosciences*, vol. 23, no. 1, pp. 1–9, 2024. doi: 10.1016/j.acags.2024.100178.