Lightweight Hybrid Feature Engineering and Fusion with Variational Autoencoder and XGBoost for Nutritional Data Classification

Wiwien Hadikurniawati
Faculty of Information Technology, Satya Wacana Christian University, Salatiga 50715, Indonesia | Faculty of Information Technology and Industry, Universitas Stikubank, Semarang 50241, Indonesia
982022007@student.uksw.edu (corresponding author)

Kristoko Dwi Hartomo
Faculty of Information Technology, Satya Wacana Christian University, Salatiga 50715, Indonesia
kristoko@uksw.edu

Irwan Sembiring
Faculty of Information Technology, Satya Wacana Christian University, Salatiga 50715, Indonesia
irwan@uksw.edu

Received: 4 October 2025 | Revised: 25 October 2025 and 30 October 2025 | Accepted: 3 November 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.15310*

ABSTRACT

Stunting remains a significant global health challenge, particularly among children under five years old. Early detection of nutritional status is crucial to prevent long-term impacts on both physical and cognitive development. This study proposes a lightweight approach based on hybrid feature engineering and fusion for stunting status classification without requiring complex data balancing strategies. The proposed method integrates basic anthropometric features with Rule-Based (RB) derived features (BMI, weight, and height changes since birth, growth rates, and z-score combinations), along with latent features generated by a Variational Autoencoder (VAE) that captures non-linear patterns in tabular data. All features are then concatenated and classified using XGBoost. The evaluation with 5-fold cross-validation demonstrated highly competitive performance, yielding an average F1-score of 0.9919 ± 0.0066 on the testing data. These findings highlight that a lightweight yet informative approach can serve as a practical solution to support early stunting detection, particularly in public health settings with limited resources.

Keywords-stunting prediction; feature engineering; Variational Autoencoder; XGBoost; lightweight model

I. INTRODUCTION

Stunting is a chronic nutritional problem that remains a serious concern in global public health. According to UNICEF, WHO, and the World Bank (2023), approximately 22.3% of children under five worldwide experience stunting [1]. This figure is particularly striking in South Asia (30.5%), Sub-Saharan Africa (31.3%), Southeast Asia (26.4%), and Oceania (excluding Australia and New Zealand) (44.0%), where economic factors, food access, and health services remain major obstacles [2-4]. In Indonesia, the prevalence of stunting among children remained at 21.6% in 2022 according to the Indonesian Nutritional Status Survey (SSGI) [5]. This situation demands an accurate, efficient early-detection strategy to support community nutrition interventions.

Traditional approaches to determining children's nutritional status generally rely on anthropometric parameters, such as weight, height, and z-scores, derived from WHO standards. Although simple and easy to measure, these methods are static and less adaptable in capturing the complex patterns of child growth interactions [6]. Therefore, some studies have begun to adopt ML to improve the accuracy of nutritional status classification [7-9]. ML models, such as Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression, often outperform conventional statistical methods in many empirical studies, especially for classification or prediction accuracy on complex datasets. However, they remain vulnerable to overfitting and limited generalization, especially when the dataset is small, features are numerous or unclean, class distribution is imbalanced, or when model validation and tuning procedures are inadequate [10-13]. On the other hand, advances in deep learning (DL) have introduced models, such as Convolutional Neural Networks (CNN) and TabNet, which are capable of extracting richer feature representations [14-16]. However, these architectures are computationally intensive and often require additional strategies such as balancing techniques

[17, 18]. This limits the applicability of the models to real-world environments, particularly primary healthcare facilities in resource-constrained areas.

To address these challenges, this study proposes a lightweight hybrid feature engineering and fusion approach [19, 20]. This method combines basic anthropometric features with RB derived features (such as body mass index, weight/height change since birth, and growth rate) and latent features from a VAE capable of capturing non-linear representations of tabular data [21, 22]. All features are then combined through a simple fusion mechanism and processed using XGBoost, a boosting-based tree algorithm that is efficient, stable, and proven robust on tabular data [23-25]. The main contributions of this study are:

- Developing a robust and lightweight pipeline without the need for data balancing.
- Integrating RB interpretable features with latent features from VAE, enabling the model to capture both classical patterns and non-linear representations.
- Demonstrating the effectiveness of XGBoost as a stable, efficient, and robust primary classification model for stunting prediction.

II. RELATED WORKS

Extensive research on ML-based stunting prediction has been conducted to improve the accuracy of early detection compared to conventional statistical methods. Authors in [9] used data from the 2016 Ethiopia Demographic and Health Survey (EDHS) covering over 9,000 children under five to evaluate several algorithms, including Logistic Regression, Decision Tree, RF, SVM, and XGBoost. The results showed that RF and XGBoost provided accuracy above 80%, with child age, maternal education, and family economic status as the main determinants. However, this study did not address data imbalance and used only standard tabular features, limiting its ability to capture more complex nonlinear patterns.

In addition to focusing on toddlers, authors in [8] examined the risk of stunting in adolescent girls aged 15–19 years using 2016 EDHS data from a sample of 3,156. This study compared eight ML algorithms (Logistic Regression, RF, SVM, KNN, Naive Bayes, Decision Tree, XGBoost, and LightGBM), and applied SMOTE for data balancing and Boruta for feature selection. The results showed that RF was again the best model with an AUC of 95%, and region, age, wealth index, education, and sanitation facilities emerged as significant predictors. Although the analysis was enriched with association rule mining and interpretability through SHAP values, this study was still limited to tabular survey data without clinical nutrition information or longitudinal data. Thus, the generalizability of the results remains low.

To broaden the scope of the analysis, authors in [26] combined 2011 and 2016 EDHS data with over 18,000 toddler samples to develop a multi-class classification model that differentiates stunting status into normal, moderate, and severe. After preprocessing, feature selection, and balancing with SMOTE, four ensemble algorithms were compared: RF, AdaBoost, XGBoost, and CatBoost. The results showed that RF provided the highest performance with an accuracy approaching 98% and an AUC of 99.99%. It successfully identified child age, maternal education, anemia, wealth index, and access to clean water and sanitation as key determinants.

Several studies also highlight the importance of model interpretability. Authors in [27] used BDHS 2017–18 data from Bangladesh and evaluated Logistic Regression, ANN, RF, and XGBoost. XGBoost emerged as the best model with an accuracy of 81.73% and an AUC of 0.802, while SHAP was used to identify key factors such as parental education, economic status, BMI, and access to sanitation. In Indonesia, authors in [28] introduced a hybrid approach combining classification (SVM with RBF and Sigmoid kernels), linear regression for trend prediction, and K-Medoids-based spatial clustering. The results showed 91.3% accuracy for classification and a low MSE for regression, while providing a spatial overview of stunting prevalence at the regional level.

Furthermore, authors in [14] proposed a hybrid dual-fusion framework that combines deep features from a 1D-CNN and classical anthropometric indicators via an additive attention mechanism, and then classifies them using a stacking ensemble. To address data imbalance, SMOTE was applied in the present study and achieved very high performance (99.7% accuracy and 99.99% AUC). However, this approach still suffers from limitations, including relatively high architectural complexity, the need for external balancing, and high computational costs, making its application in primary healthcare less practical.

It can be observed that most studies still rely on external balancing or complex models to achieve high performance. At the same time, the integration of lightweight yet informative feature engineering has not been widely explored. Therefore, the current study offers a distinct contribution by developing a robust, lightweight pipeline without balancing, integrating rule-based, interpretable features with latent features from VAEs, and using XGBoost as an efficient, stable primary model for stunting prediction in real-world, resource-constrained environments.

In addition, in the context of malnutrition prediction, few studies have incorporated explainable AI (XAI) methods, such as SHAP, to improve model interpretability. Research in the broader medical and healthcare domain has demonstrated that SHAP-based explanations for feature contributions can improve clinical trust and adoption, particularly in data-limited or resource-constrained environments [29].

III. PROPOSED METHOD

The proposed method aims to develop a lightweight yet effective hybrid pipeline for predicting toddler nutritional status using basic anthropometric data. The overall workflow consists of four main stages: data preprocessing, feature engineering (RB and VAE-based), feature fusion with re-normalization, and classification using XGBoost. The entire process, illustrated in Figure 1, is evaluated using 5-fold stratified cross-validation to ensure robust, unbiased performance estimation.

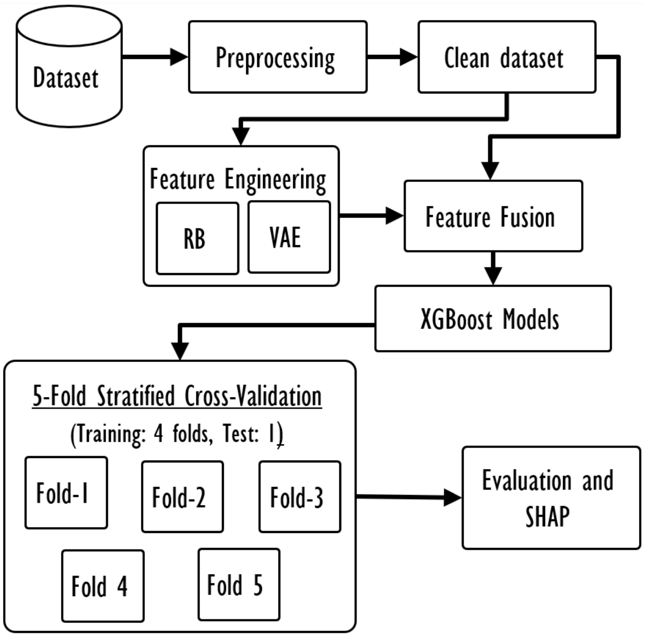

Fig. 1. Workflow of the proposed hybrid pipeline, including data preprocessing, RB and VAE feature engineering, feature fusion, and evaluation through 5-fold stratified cross-validation.

A. Preprocessing

The dataset used in this study was obtained from the Semarang City Health Office, Indonesia. This dataset has the following features: age in months (*A*), birth weight (*BW*), birth height (*BH*), weight at measurement (*W*), height at measurement (*H*), weight-for-age z-score (ZS_{WA}), height-for-age z-score (ZS_{HA}), weight-for-height z-score (ZS_{WH}), and the child's age in months. The labels used include stunting status and nutritional status. Next, this dataset was preprocessed with the following steps:

- Removal of irrelevant features, such as administrative columns (Parent Name, Health Post, Neighborhood (RT-RW), Date of Birth, and Measurement Date) were excluded. In addition, the Mid-Upper Arm Circumference (MUAC) feature was removed because a large proportion of its values were recorded as zero, likely representing missing or erroneous entries.
- Records containing missing values or duplicates were removed to maintain data consistency.
- Categorical variables (e.g., sex) were converted into numerical representations using one-hot encoding.
- All numerical features were normalized utilizing z-score scaling, using:

$$x' = \frac{x-\mu}{\sigma} \tag{1}$$

where μ is the mean and σ is the standard deviation of each feature. This ensures that the feature distributions are comparable in scale to support representation learning. The 5-fold cross-validation was applied on the training set to obtain a robust internal estimate of model performance to ensure an unbiased measure of generalization.

To strictly prevent data leakage, all preprocessing operations were restricted to the training data. Specifically, the z-score normalization in (1) was fit exclusively to the training subset and then applied to the validation and test subsets. During each cross-validation fold, the scaler was re-initialized and re-fitted using only the samples in the corresponding training fold, ensuring that no statistical information from the validation or test data influenced the feature scaling.

B. Feature Engineering

To improve data representation, two main approaches were used: RB features and feature transformation through a VAE. RB features were derived from original anthropometric features to capture child growth dynamics, such as body mass index (BMI), weight difference from birth (ΔW), height difference from birth (ΔH), weight growth rate GR_W, height growth rate GR_H, mean of z-scores, and difference of z-scores. Some RB features were inspired by [30-33], where nutritional calculations were discussed. The formulas used are:

$$BMI = \frac{W}{H^2} \tag{2}$$

$$\Delta W = W - BW \tag{3}$$

$$\Delta H = H - BH \tag{4}$$

$$GR_W = \frac{\Delta W}{A} \tag{5}$$

$$GR_H = \frac{\Delta H}{A} \tag{6}$$

$$ZS_{mean} = \frac{ZS_{WA}+ZS_{HA}+ZS_{WH}}{3} \tag{7}$$

$$ZS_{diff} = ZS_{WA} - ZS_{HA} \tag{8}$$

RB features add clinical context, such as whether weight gain follows a normal rate with age, or whether there is an imbalance between weight and height based on WHO standards. VAE also includes features. Unlike simple feature selection, VAE performs feature transformation from the original input space into a low-dimensional latent representation [34, 35]. The VAE steps are:

- All original numerical features are used (BW, BH, W, H, ZS_{WA}, ZS_{HA}, ZS_{WH}, A) as input.
- The output is a compact latent vector ($z \in \mathbb{R}^4$) representing the non-linear pattern between anthropometric features.

In principle, the encoder maps the input into a parameterized Gaussian distribution (μ, σ) from which a latent vector z is generated through the reparameterization trick. The decoder is used only during training for reconstruction, ensuring that the learned representation remains informative; during inference, only the encoder is used [36]. Consequently, the VAE serves to generate new features (VAE_{z1}–VAE_{z4}) that are more compact and capable of capturing complex interactions among anthropometric variables, thereby enriching the dataset representation before being combined in the feature fusion stage. The configuration is presented in Table I. The

VAE model was trained solely on the data from each split, and the trained encoder was subsequently used to extract latent representations for both the training and test subsets, thereby maintaining complete isolation between the training and evaluation data.

TABLE I. VAE CONFIGURATION

Component	Architecture Design	Notes
Encoder	Dense (18) → Dense (12) → Dense (6)	Input: 8 dims; Activation: ReLU
Latent Heads	Dense (4) for μ and Dense (4) for log σ² (two parallel heads)	Activation: Linear
Decoder	Dense (6) → Dense (12) → Dense (8)	Output: 8 dims (reconstruct input)
Regularization	KL divergence with β = 0.5	β is empirically chosen for stable training and to avoid over-regularization
Optimizer	Adam (*lr* = 1e-3)	Batch size = 64, epochs = 50, early stopping (patience 7)

Table II presents the features used in this study after preprocessing and feature engineering.

TABLE II. SUMMARY OF FEATURES USED IN THE STUDY

Feature Name	Type	Description
Birth Weight (BW)	Original	Weight of the child at birth (kg)
Birth Height (BH)	Original	Length/height of the child at birth (cm)
Current Weight (W)	Original	Current body weight (kg)
Current Height (H)	Original	Current body height (cm)
Age (A)	Original	Age in months, calculated from measurement date – birth date
z-score W/A (ZS_{WA})	Original	WHO standardized z-score of weight-for-age
z-score H/A (ZS_{HA})	Original	WHO standardized z-score of height-for-age
z-score W/H (ZS_{WH})	Original	WHO standardized z-score of weight-for-height
BMI	Engineered	Body Mass Index = Weight / Height²
ΔW	Engineered	The difference between the current weight and the birth weight
ΔH	Engineered	Difference between current height and birth height
Growth Rate (GR_W)	Engineered	$\Delta W/A$
Growth Rate (GR_H)	Engineered	$\Delta H/A$
ZS_{mean}	Engineered	Mean of ZS_{WA}, ZS_{HA}, ZS_{WH}
ZS_{diff}	Engineered	Difference between ZS_{WA} and ZS_{HA}
(VAE_{z1}–VAE_{z4})	Engineered	Latent variables learned by VAE
Status_Nutrition	Label	multi-class label of nutritional status

C. Feature Fusion

There are three sets of original features before reduction: nine original features, seven RB features, and four latent features, for a total of 19. All features are concatenated, as shown in:

$$F_{\text{fusion}} = [F_{\text{ori}}||F_{\text{rule}}||F_{\text{latent}}] \in \mathbb{R}^{19} \quad (9)$$

Subsequently, re-normalization using z-score scaling is applied to F_{fusion} to standardize the mixed scales (original, derived, and latent features). This step is important because $VAE_z \sim N(0,1)$, but its combination with RB features often alters the overall distribution.

D. Classification Model

The main algorithm is XGBoost, a gradient-boosting Decision Tree with regularization to reduce overfitting. The XGBoost objective function is defined as in:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (10)$$

where l denotes the loss function (logistic loss for binary classification) and Ω represents the regularization term of tree complexity. The XGBoost configuration used in this study is summarized in Table III.

TABLE III. XGBOOST PARAMETER SETTINGS

Parameter	Value	Parameters	Value
learning_rate	0.05	colsample_bytree	0.8
max_depth	6	gamma	0.1
n_estimators	300	reg_alpha (L_1)	0.1
subsample	0.8	reg_lambda (L_2)	1.0

The dataset was divided into 80% training data and 20% testing data, stratified by class label to maintain the distribution proportions. Next, the training data were evaluated using 5-fold cross-validation, with each fold serving as validation data, while the remainder was used to train the model. This process yielded more robust performance estimates and reduced the bias caused by single data splits. After cross-validation, the final model was retrained on the entire training data (80%) and evaluated on the testing data (20%). The evaluation used metrics, such as accuracy, precision, recall, F1-score, multiclass AUC, and confusion matrix analysis, to measure overall classification performance.

The proposed approach is not limited to the Indonesian context. Since the method primarily relies on basic anthropometric indicators standardized by WHO (such as weight, height, age, and z-scores), the same pipeline can be applied to similar datasets from other countries. This highlights the adaptability of the proposed lightweight hybrid design to broader public health applications in low-resource settings.

E. Evaluation and Explainable AI

Model performance was evaluated using five standard classification metrics: accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Accuracy reflects the overall correctness of predictions, while precision and recall measure the model's ability to identify positive cases and avoid false negatives, respectively. The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation when the dataset is imbalanced. Additionally, the AUC metric was used to assess the model's discriminative capability across all classification thresholds.

Beyond quantitative evaluation, XAI analysis was conducted to interpret how each feature contributes to the model's decision process. The SHapley Additive exPlanations (SHAP) framework was employed to quantify the contribution of each input variable to the predicted outcomes. Global feature importance (mean |SHAP| values) were generated to identify the most influential features. This integration of performance evaluation and explainability provides not only a robust assessment of predictive accuracy but also valuable insights into the underlying feature interactions relevant to nutritional status prediction.

IV. RESULTS AND DISCUSSION

A. Data Distribution

The dataset used in this study consists of 2,752 child growth records collected from the Semarang City Health Office (Indonesia). As shown in Figure 2, the dataset is highly imbalanced, with most samples belonging to the Well-Nutrition class (n = 2,344). The remaining classes include At Risk of Overweight (n = 218), Underweight (n = 88), Overweight (n = 67), Obesity (n = 25), and Malnutrition (n = 10). This imbalance poses a potential challenge for the model to recognize minority classes, such as Malnutrition and Obesity, as illustrated in Figure 2. To ensure fair evaluation, the dataset was randomly split into 80/20 stratified sets, yielding 2,201 training and 551 testing samples, preserving the class proportions in both subsets. The stratification ensures that each class in the test data reflects approximately the same relative frequency as in the overall dataset, preventing bias toward dominant classes.

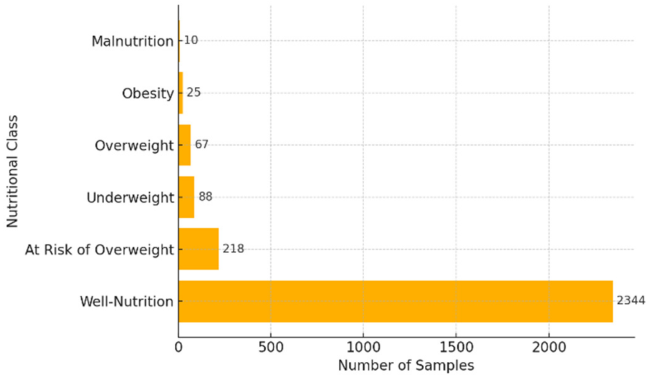

Fig. 2. Class distribution of the nutritional dataset used.

B. Cross-Fold Validation

During the training process, an evaluation was conducted using 5-fold stratified cross-validation on the training and testing data. The results per fold showed consistent high performance, with a weighted F1-score ranging from 0.9822 to 1.0000. The average F1-weighted score reached 0.9919 ± 0.0066, indicating model stability across various subsets of the training data. This demonstrates that the developed pipeline is capable of addressing distribution variations between folds despite class imbalance. More detailed data are presented in Table IV. The ROC plot is displayed in Figure 3.

The proposed model accurately predicted all nutritional status categories, including minority classes such as Malnutrition and Obesity, achieving perfect classification in these groups. These results confirm the model's strong capability for nutritional status prediction under severe class imbalance. The aggregated confusion matrix of all folds, portrayed in Figure 4, further illustrates that nearly all instances were correctly classified across categories. For brevity, per-class performance is summarized in Table IV using the weighted F1-scores, which inherently account for class imbalance across folds.

TABLE IV. PROPOSED METHOD RESULTS

Fold	Accuracy	Precision	Recall	F1-score
1	0.9955	0.9955	0.9955	0.9954
2	1.0000	1.0000	1.0000	1.0000
3	0.9841	0.9812	0.9841	0.9822
4	0.9864	0.9872	0.9864	0.9863
5	0.9955	0.9957	0.9955	0.9955
Avg	0.9923 ± 0.0060	0.9919 ± 0.0068	0.9923 ± 0.0060	0.9919 ± 0.0066

C. Discussion

The test results of the proposed method highlight three key findings. First, robustness to imbalanced data. The absence of misclassifications in minority classes (e.g., Malnutrition and Obesity) confirms that the pipeline can generalize well even without additional balancing strategies such as random oversampling, SMOTE, or ADASYN.

Second, the effectiveness of feature combinations. Anthropometric-derived features (BMI, ΔW, ΔH, growth rate, etc.) provide a meaningful clinical context, while VAE-derived latent representations capture complex non-linear interactions among anthropometric indicators. The fusion of explicit domain knowledge and compact learned features makes XGBoost more efficient and stable in tabular learning tasks.

Third, regarding generalization and potential overfitting, the 5-fold cross-validation results (F1-score = 0.9919 ± 0.0066) show strong consistency across folds, indicating a minimal generalization gap. The aggregated confusion matrix in Figure 3 further supports this observation, showing that almost all classes, including minority ones, were correctly classified. Moreover, the ablation study demonstrates that each component (original, RB, and VAE features) contributes incrementally to the overall performance, reinforcing that the model captures relevant patterns rather than overfitting to the dataset.

Robustness was also implicitly evaluated through 5-fold stratified cross-validation, where the consistent performance across folds indicates strong model stability despite class imbalance. In addition, the model's resilience to data variability suggests potential robustness against moderate noise or perturbations. Nevertheless, future validation on larger or external datasets is proposed to further assess cross-population generalization.

D. Ablation Studies and Comparison

To understand the contribution of each component in the system design, an ablation study was conducted focusing on variations of architecture and features. Five testing scenarios were prepared:

- A0 (Original feature only): using only eight original anthropometric features without adding derived or latent features.
- A1 (Original + RB): using the original features plus seven RB derived features, including body mass index, weight and height differences, growth rates, and z-score summaries.
- A2 (Original + VAE): using the eight original features combined with four latent features generated by the VAE transformation.
- A3 (Fusion: Original + RB + VAE): integrating all features (8 original + 7 RB + 4 latent) into a total of 19 features, followed by re-normalization to balance the scales. This scenario represents the main design.
- A4 (Fusion without Re-normalization): identical to A3 but without re-normalization after feature fusion, to evaluate the impact of re-scaling.

In addition to the ablation study, the study also conducted a comparison with related works that utilized the same dataset. Table V provides detailed cross-validation results of the ablation and comparison experiments.

TABLE V. ABLATION STUDY AND COMPARISON RESULTS

Model	**Acc**	**Prec**	**Recall**	**F1-score**	**ROC**
A0	0.982	0.982	0.981	0.981	0.995
A1	0.989	0.989	0.988	0.988	0.997
A2	0.986	0.985	0.985	0.985	0.996
A3	0.992	0.992	0.992	0.992	1.000
A4	0.983	0.983	0.982	0.982	0.995
CNN–SMOTE [14]	0.997	0.997	0.997	0.967	0.999

Table V presents the results of ablation experiments evaluating the contribution of each feature component to the system design. The pattern remained consistent: the baseline model A0, which uses only raw anthropometric features, performed the lowest in cross-validation (Accuracy and F1 ≈ 0.981). This is understandable, as the original variables tend to be static and less able to capture the dynamics of child growth. When RB features were added (A1), accuracy and F1 improved significantly (≈0.992). Simple clinical indicators, such as BMI, birth weight/height difference, and growth rate proved to provide additional information that is highly medically relevant. This performance increase indicates that domain knowledge does play a significant role.

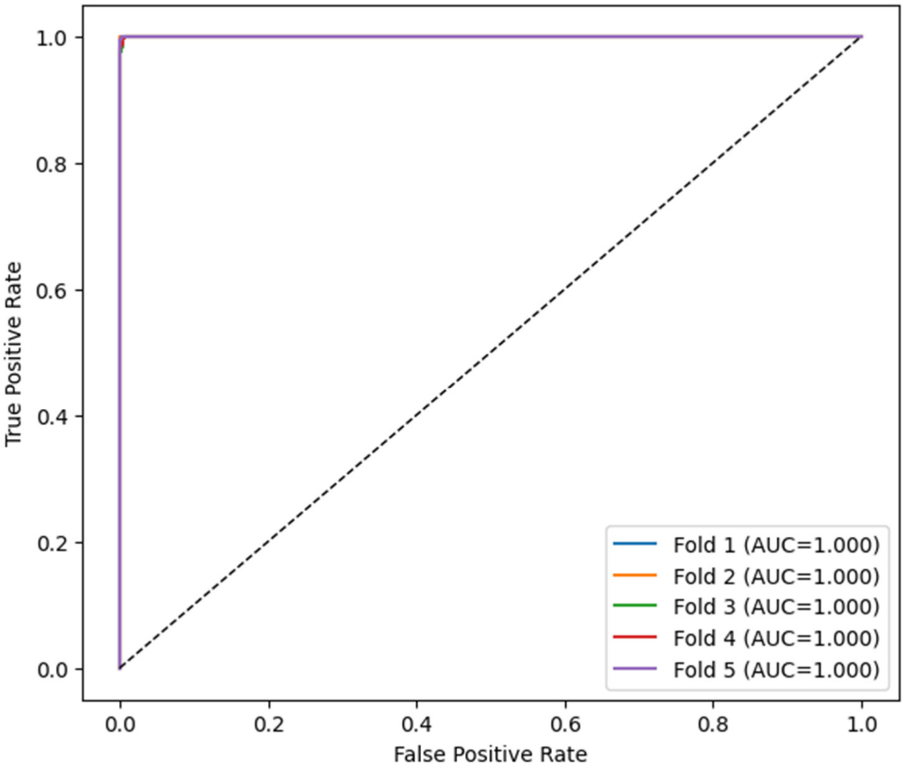

Fig. 3. ROC–AUC curves from 5-fold cross-validation showing near-perfect separability and consistent model performance across folds.

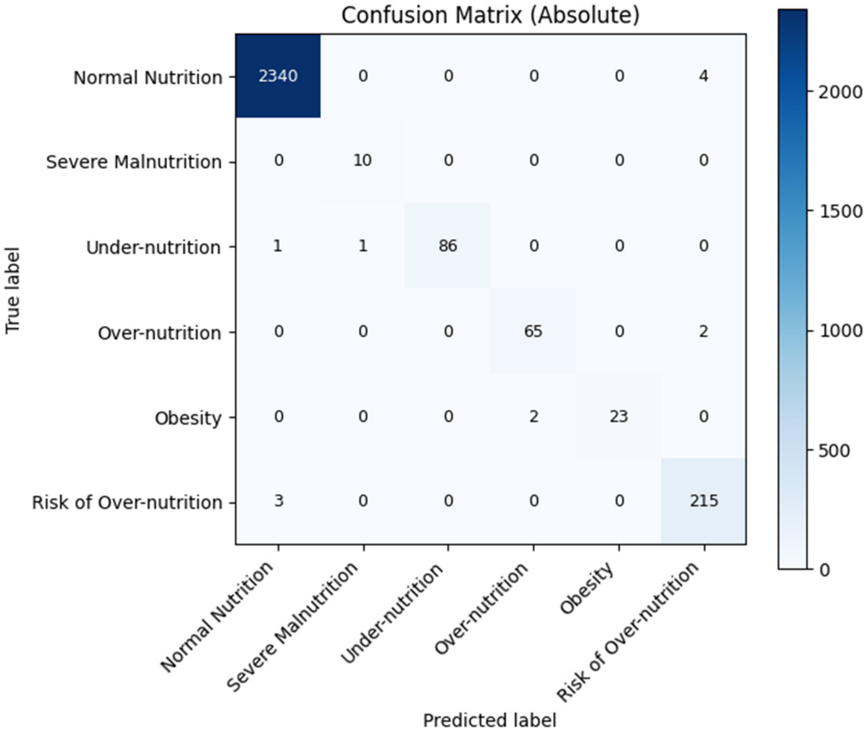

Fig. 4. Aggregated absolute confusion matrix from 5-fold cross-validation, illustrating correct classification across all six nutritional categories, including minority classes.

In A2, which relies on latent representations from VAEs, performance also improved compared to A0 (F1 ≈ 0.985). Although the improvement was slightly lower than that of A1, it indicates that nonlinear transformations can capture complex patterns among variables, even without explicitly conveying clinical significance. The results for A1 and A2 are relatively close, suggesting that both RB and VAE add value, but in different ways: one based on explicit clinical knowledge and the other on mathematical abstraction.

The combination of the two in A3 (RB + VAE) yielded the highest performance (F1 ≈ 0.992 in 5-fold cross-validation). This confirms the initial hypothesis: explicit clinical features and latent representations are complementary. RB helps the model capture familiar growth patterns, while VAE provides a more subtle non-linear representation that is difficult to capture with manually derived features. This synergy between interpretable RB and learned latent features results in a robust yet efficient configuration. When compared with existing approaches, the proposed hybrid model combining RB and VAE features with XGBoost achieved comparable or even higher accuracy than more complex DL architectures such as CNN–SMOTE ensembles [14]. This demonstrates that the proposed lightweight pipeline can deliver state-of-the-art performance without artificial data balancing or heavy computation. Furthermore, the novelty of this work lies in the integration of interpretable RB features and compact VAE representations within a single, resource-efficient framework, which has not been previously applied to nutritional status classification.

Experiment A4, without re-normalization after fusion, showed a slight decrease (F1 ≈ 0.982). While the decrease was not drastic, this result demonstrates the importance of maintaining balanced feature scales. Without normalization, the contributions of RB and VAE are not fully optimized, as the data distribution becomes more variable.

These results reveal that each component of the pipeline makes a significant contribution. The full pipeline (A3) proved most effective, while A0 served as a baseline, demonstrating the limitations of the raw features. Thus, the research hypothesis—that the integration of RB and VAE can improve the robustness of toddler nutritional status prediction—is supported by the cross-validation results. Compared with [14], where a CNN ensemble and SMOTE were deployed to achieve an accuracy approaching 99.7%, this study shows that the proposed model can match or exceed that performance with a much lighter pipeline and without artificial balancing. To validate the reliability of this improvement, a Wilcoxon signed-rank test was conducted between the baseline (A0) and the proposed hybrid model (A3) using F1-scores from five cross-validation folds. The result ($p < 0.05$) confirmed that the performance gain was statistically significant rather than due to random variation, indicating that the RB + VAE integration provides a genuine and reliable enhancement over the baseline. This emphasizes that lightweight yet informative designs can be a more practical alternative for real-world implementations.

E. Explainable AI (XAI)

To further interpret the model's decision-making process, SHAP analysis was performed to identify the relative contribution of each feature to the prediction outcomes. The global feature importance, based on mean absolute SHAP values, is displayed in Figure 5. The results show that ZS_{WH} (weight-for-height z-score) emerged as the most influential factor, followed by *BMI*, ZS_{diff}, and ZS_{WA}, indicating that anthropometric indicators derived from weight and height had the greatest impact on nutritional status classification.

Features, such as ZS_{mean}, GR_W, and W, contributed moderately, while variables, like dW, ZS_{HA}, and A, exhibited lower SHAP values, suggesting a smaller effect on model predictions. The dominance of ZS_{WH} and *BMI* aligns with the clinical understanding that both indices are sensitive to acute and chronic nutritional imbalances. This consistency between model interpretability and physiological relevance demonstrates that the proposed lightweight model not only achieves high predictive accuracy but also produces explainable and trustworthy predictions in the context of child nutritional assessment.

The proposed lightweight and interpretable model has potential for real-world adoption in early stunting and nutritional risk detection programs. Its low computational complexity and explainable structure make it suitable for deployment at community health centers (Posyandu) or within national child nutrition monitoring systems. By integrating the model into routine anthropometric data collection, local health workers could obtain immediate, automated screening results with interpretable indicators such as BMI and weight-for-height z-scores. This would enable early identification of at-risk children and support timely interventions without relying on complex laboratory analyses or high-end computational infrastructure. Overall, the presented results—including the cross-validation metrics in Tables IV and V, ROC and confusion matrix in Figures 2 and 3, and SHAP-based feature interpretability—confirm the claims made by the study regarding model robustness and high predictive performance.

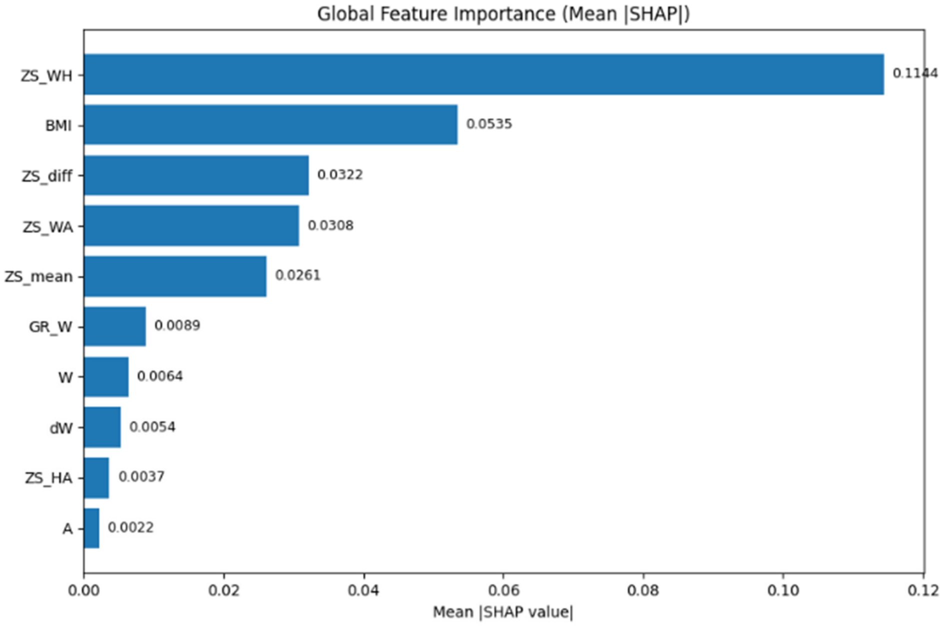

Fig. 5. Mean absolute SHAP values representing the global importance of features in the nutritional status prediction model.

V. CONCLUSIONS

The present research efficiently developed a lightweight hybrid pipeline for predicting toddler nutritional status. It combined basic anthropometric features, Rule-Based (RB) interpretable features, and latent representations from Variational Autoencoders (VAEs), which were then fused and processed using XGBoost. The evaluation results based on 5-fold stratified cross-validation show that this pipeline achieves very high, stable performance (average F1-score = 0.9919 ± 0.0066, and AUC ≈ 1.000), even without data balancing techniques such as SMOTE.

Beyond performance, the main contribution of this study is an efficient, interpretable, and resource-friendly approach that can be feasibly implemented in community health services with limited infrastructure. By relying solely on simple anthropometric inputs, enriched with engineered and latent features, the method provides a practical solution that balances computational efficiency with clinical relevance.

Nevertheless, several limitations need to be considered. The dataset used is limited to a single region, so further validation on multi-regional or publicly available datasets is required to confirm the model's widespread applicability. In addition, while the current results show no signs of overfitting, broader testing is essential to ensure robustness across diverse populations.

For research can be extended toward: (i) cross-regional and public dataset validation, (ii) applying model interpretability techniques to strengthen clinical decision support, and (iii) developing mobile or web-based tools. So, the proposed pipeline can be directly adopted by public health services, especially in resource-constrained settings.

ETHICS AND DATA AVAILABILITY STATEMENT

The dataset used in this study was obtained with permission from the Semarang City Health Office, Indonesia, and contains anonymized anthropometric records of toddlers. All data were processed in compliance with local health data regulations and used solely for research purposes without any personally identifiable information. Due to privacy and data-sharing restrictions, the dataset is not publicly available; however, it can be provided upon reasonable request to the corresponding author for academic research purposes.

REFERENCES

[1] *Levels and Trends in Child Malnutrition Child Malnutrition: Key Findings of the 2023*, 1st ed. Geneva, Switzerland: World Health Organization, 2023.

[2] *The State of Food Security and Nutrition in the World 2023. Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum*. Rome, Italy: United Nations International Children's Emergency Fund, United Nations, 2023.

[3] S. Habimana and E. Biracyaza, "Risk Factors of Stunting Among Children Under 5 Years of Age in the Eastern and Western Provinces of Rwanda: Analysis of Rwanda Demographic and Health Survey 2014/2015," *Pediatric Health, Medicine and Therapeutics*, vol. 10, pp. 115–130, Oct. 2019, https://doi.org/10.2147/PHMT.S222198.

[4] T. Vaivada, N. Akseer, S. Akseer, A. Somaskandan, M. Stefopulos, and Z. A. Bhutta, " Stunting in Childhood: An Overview of Global Burden, Trends, Determinants, and Drivers of Decline," *The American Journal of Clinical Nutrition*, vol. 112, pp. 777S-791S, Sept. 2020, https://doi.org/10.1093/ajcn/nqaa159.

[5] *Buku Saku Hasil Survei Status Gizi Indonesia (SSGI)*. Jakarta, Indonesia: Badan Kebijakan Pembangunan Kesehatan, Kementerian Kesehatan Ri, 2022.

[6] H. D. S. Ferreira, "Anthropometric Assessment of Children's Nutritional Status: A New Approach Based on an Adaptation of Waterlow's Classification," *BMC Pediatrics*, vol. 20, no. 1, Dec. 2020, Art. no. 65, https://doi.org/10.1186/s12887-020-1940-6.

[7] H. M. Fenta, T. Zewotir, and E. K. Muluneh, " A Machine Learning Classifier Approach for Identifying the Determinants of Under-five Child Undernutrition in Ethiopian Administrative Zones," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, Oct. 2021, Art. no. 291, https://doi.org/10.1186/s12911-021-01652-1.

[8] A. B. Zemariam *et al.*, " Prediction of Stunting and Its Socioeconomic Determinants Among Adolescent Girls in Ethiopia using Machine Learning Algorithms," *PLOS ONE*, vol. 20, no. 1, Jan. 2025, Art. no. e0316452, https://doi.org/10.1371/journal.pone.0316452.

[9] 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, Sept. 2023, Art. no. 1638, https://doi.org/10.3390/children10101638.

[10] D. Priyanto, H. Hairani, K. Marzuki, and M. Innuddin, "Optimization of Random Forest for Health Data Classification using PCA and K-Means SMOTE-ENN," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27646–27652, Oct. 2025, https://doi.org/10.48084/etasr.12976.

[11] A. Abu-Errub, "Improving Early Autism Detection with Chi-Square Feature Selection, Machine Learning, and Explainable AI," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 27954–27959, Oct. 2025, https://doi.org/10.48084/etasr.12999.

[12] M. Al-Duais *et al.*, "Comparative Analysis of Machine Learning and Deep learning Techniques for Early Prediction of Breast Cancer," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 2, pp. 242–254, June 2025, https://doi.org/10.62411/faith.3048-3719-68.

[13] F. Mustofa, A. N. Safriandono, A. R. Muslikh, and D. R. I. M. Setiadi, "Dataset and Feature Analysis for Diabetes Mellitus Classification using Random Forest," *Journal of Computing Theories and Applications*, vol. 1, no. 1, pp. 41–48, Sept. 2023, https://doi.org/10.33633/jcta.v1i1.9190.

[14] W. Hadikurniawati, K. D. Hartomo, I. Sembiring, and C. Arthur, "A Dual-Fusion Hybrid Model with Attention for Stunting Prediction among Children under Five Years," *Journal of Applied Data Sciences*, vol. 6, no. 3, pp. 1985–1998, Sept. 2025.

[15] Y. Jin *et al.*, "Classification of Alzheimer's Disease using Robust TabNet Neural Networks on Genetic Data," *Mathematical Biosciences and Engineering*, vol. 20, no. 5, pp. 8358–8374, 2023, https://doi.org/10.3934/mbe.2023366.

[16] K. Pyar, "Integrating Convolutional Neural Network and Weighted Moving Average for Enhanced Human Fall Detection Performance," *Journal of Computing Theories and Applications*, vol. 2, no. 1, pp. 13–21, May 2024, https://doi.org/10.62411/jcta.10428.

[17] T. Sugihartono, "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, Jan. 2024, https://doi.org/10.47738/jads.v6i1.494.

[18] S. Hussain *et al.*, "An Enhanced Random Forest (ERF)-based Machine Learning Framework for Resampling, Prediction, and Classification of Mobile Applications using Textual Features," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19776–19781, Feb. 2025, https://doi.org/10.48084/etasr.9148.

[19] M. G. S. P. Kusuma *et al.*, "Integrating Quantum, Deep, and Classic Features with Attention-Guided AdaBoost for Medical Risk Prediction," *Journal of Computing Theories and Applications*, vol. 3, no. 2, pp. 171–189, Oct. 2025, https://doi.org/10.62411/jcta.14873.

[20] M. N. Aisy, S. A. Wulandari, and D. R. I. M. Setiadi, "A Probabilistic Feature-Augmented GRU-Attention Model for Chronic Disease Prediction on Imbalanced Data," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 2, pp. 282–293, July 2025, https://doi.org/10.62411/faith.3048-3719-100.

[21] F. Ferré, S. Allassonnière, C. Chadebec, and V. Minville, "Generating Artificial Patients with Reliable Clinical Characteristics using a Geometry-Based Variational Autoencoder: Proof-of-Concept Feasibility Study," *Journal of Medical Internet Research*, vol. 27, Apr. 2025, Art. no. e63130, https://doi.org/10.2196/63130.

[22] N. Simidjievski *et al.*, "Variational Autoencoders for Cancer Data Integration: Design Principles and Computational Practice," *Frontiers in Genetics*, vol. 10, Dec. 2019, Art. no. 1205, https://doi.org/10.3389/fgene.2019.01205.

[23] C. C. Odiakaose *et al.*, "Hypertension Detection via Tree-Based Stack Ensemble with SMOTE-Tomek Data Balance and XGBoost Meta-Learner," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 3, pp. 269–283, Dec. 2024, https://doi.org/10.62411/faith.3048-3719-43.

[24] D. R. I. M. Setiadi, K. Nugroho, A. R. Muslikh, S. W. Iriananda, and A. A. Ojugo, "Integrating SMOTE-Tomek and Fusion Learning with XGBoost Meta-Learner for Robust Diabetes Recognition," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 1, pp. 23–38, May 2024, https://doi.org/10.62411/faith.2024-11.

[25] R. Kansal and C. Diwaker, "Efficiency Determination of Various Machine Learning Techniques for Sentiment Analysis on Social Media Platforms," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25584–25589, Aug. 2025, https://doi.org/10.48084/etasr.11158.

[26] M. K. Ayele, G. A. Baye, S. H. Yesuf, A. A. Engda, and E. T. Mitiku, " Predicting Stunting Status Among Under Five Children in Ethiopia using Ensemblemachine Learning Algorithms," *Scientific Reports*, vol. 15, no. 1, July 2025, Art. no. 27907, https://doi.org/10.1038/s41598-025-03206-1.

[27] Md. M. Islam, N. Md. S. J. Kibria, S. Kumar, D. C. Roy, and Md. R. Karim, " Prediction of Undernutrition and Identification of its Influencing Predictors Among Under-five Children in Bangladesh using Explainable Machine Learning Algorithms," *PLOS ONE*, vol. 19, no. 12,

Dec. 2024, Art. no. e0315393, https://doi.org/10.1371/journal.pone.0315393.

[28] N. Hasdyna, R. K. Dinata, Rahmi, and T. I. Fajri, "Hybrid Machine Learning for Stunting Prevalence: A Novel Comprehensive Approach to Its Classification, Prediction, and Clustering Optimization in Aceh, Indonesia," *Informatics*, vol. 11, no. 4, Nov. 2024, Art. no. 89, https://doi.org/10.3390/informatics11040089.

[29] T. R. Noviandy, G. M. Idroes, and I. Hardi, "An Interpretable Machine Learning Strategy for Antimalarial Drug Discovery with LightGBM and SHAP," *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 2, pp. 84–95, Aug. 2024, https://doi.org/10.62411/faith.2024-16.

[30] C. A. Akiotu *et al.*, "A Predictive Nutritional Assessment System for Vegetarians using Artificial Neural Networks," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 2, pp. 294–312, Aug. 2025, https://doi.org/10.62411/faith.3048-3719-117.

[31] A. Schreuder, E. Corpeleijn, and T. Vrijkotte, " Modelling Individual Infancy Growth Trajectories to Predict Excessive Gain in BMI z-score: A Comparison of Growth Measures in the Abcd and Gecko Drenthe Cohorts," *BMC Public Health*, vol. 23, no. 1, Dec. 2023, Art. no. 2428, https://doi.org/10.1186/s12889-023-17354-4.

[32] J. Mkungudza, H. S. Twabi, and S. O. M. Manda, " Development of a Diagnostic Predictive Model for Determining Child Stunting in Malawi: A Comparative Analysis of Variable Selection Approaches," *BMC Medical Research Methodology*, vol. 24, no. 1, Aug. 2024, Art. no. 175, https://doi.org/10.1186/s12874-024-02283-6.

[33] A. Hendy *et al.*, " Supervised Machine Learning for Classification and Prediction of Stunting Among Under-five Egyptian Children," *BMC Pediatrics*, vol. 25, no. 1, Sept. 2025, Art. no. 681, https://doi.org/10.1186/s12887-025-06138-x.

[34] R. Wei and A. Mahmood, "Recent Advances in Variational Autoencoders with Representation Learning for Biomedical Informatics: A Survey," *IEEE Access*, vol. 9, pp. 4939–4956, 2021, https://doi.org/10.1109/ACCESS.2020.3048309.

[35] D. P. Gomari *et al.*, "Variational Autoencoders learn transferrable representations of metabolomics data," *Communications Biology*, vol. 5, no. 1, June 2022, Art. no. 645, https://doi.org/10.1038/s42003-022-03579-3.

[36] R. Wei, C. Garcia, A. El-Sayed, V. Peterson, and A. Mahmood, "Variations in Variational Autoencoders - A Comparative Evaluation," *IEEE Access*, vol. 8, pp. 153651–153670, 2020, https://doi.org/10.1109/ACCESS.2020.3018151.