ISSN 2723-6471

A Dual-Fusion Hybrid Model with Attention for Stunting Prediction among Children under Five Years

Wiwien Hadikurniawati[1], Kristoko Dwi Hartomo[2,*], Irwan Sembiring[3], Christian Arthur[4]

[1,2,3]*Faculty of Information Technology, Satya Wacana Christian University, Salatiga 50715, Indonesia*

[1]*Faculty of Information Technology and Industry, Universitas Stikubank, Semarang 50241, Indonesia*

[4]*School of Transdiciplinary, University of Technology Sydney, Ultimo, NSW 2007, Australia*

(Received: January 10, 2025; Revised: February 12, 2025; Accepted: May 15, 2025; Available online: July 13, 2025)

Abstract

Malnutrition remains a persistent global health challenge, especially among children under five. Traditional assessment methods often rely on static anthropometric measures, which are limited in capturing complex growth patterns. This study aims to develop a robust classification model for predicting the nutritional status of children under five years old, addressing the critical public health challenge of stunting. The model contributes to the growing need for accurate, data-driven early detection systems in child health monitoring by introducing a hybrid framework that combines deep learning and classical machine learning techniques. The proposed approach integrates automatically extracted features from a One-Dimensional Convolutional Neural Network (1D-CNN) with classical anthropometric indicators. These combined features are processed through an additive attention mechanism, highlighting the most informative attributes. The attention-weighted representation is then classified using an ensemble stacking method that aggregates predictions from multiple base classifiers, including decision trees, nearest neighbor algorithms, support vector machines, etc. Synthetic Minority Over-sampling Technique (SMOTE) is applied to the training dataset to mitigate data imbalance, particularly the underrepresentation of severe and moderate malnutrition cases. The research utilizes a dataset comprising 2,789 records of children under five years old collected from community health posts in Indonesia. Data preprocessing included cleaning, normalization, and gender encoding. The model's performance was evaluated using 5-fold cross-validation and measured by accuracy, precision, recall, and area under the curve metrics. The results show that the proposed model achieved an average accuracy of 99.70% and an area under the curve of 99.99%. An ablation study further demonstrated the significant contribution of each component, feature extraction, fusion mechanism, and ensemble classifier to the final performance. This approach reveals a robust and scalable solution for early nutritional status prediction in healthcare settings.

Keywords: Nutritional Status Classification, 1D-CNN, Ensemble Learning, Additive Attention, SMOTE, Dual-Fusion, Child Malnutrition

1. Introduction

The problem of malnutrition, especially stunting, remains a global public health challenge [1], [2]. In Indonesia, the prevalence of stunting reached 36% based on the 2022 Nutritional Status Survey [3]. Java Island, with the highest number of toddlers, contributed a significant figure, with Central Java recording the highest prevalence of 20.8% [4]. Malnutrition impacts physical growth, cognitive development, and the risk of chronic diseases in the future [5]. Early detection of children's nutritional status is crucial to prevent these long-term impacts [6], [7].

Conventional nutritional status assessment generally uses anthropometric parameters such as weight, height, and body mass index. However, this approach is static and less adaptive to variations in growth between individuals. Machine Learning (ML) algorithms such as Decision Trees (DT), K-Nearest Neighbors (KNN), Naive Bayes (NB), and Support Vector Machine (SVM) have been used specifically for nutritional status classification [2], [8], [9], [10], [11], [12], [13]. Although the single model is easy to implement and fast in inference, its limitations lie in its sensitivity to noise and overfitting on complex datasets. In contrast, ensemble models such as stacking, boosting or voting can combine the strengths of several base models to improve the accuracy, stability, and generalization of predictions [14], [15],

***Corresponding author: Kristoko Dwi Hartomo (kristoko@uksw.edu)**
DOI: https://doi.org/10.47738/jads.v63.831

[16], [17], [18]. One of the important challenges in classification, specifically on health data sets such as nutritional status, is a class imbalance, where the amount of data in certain categories, such as undernutrition or overweight, is much less than the normal category. This imbalance can cause a model bias towards the majority class. To overcome this, a balancing strategy is needed for the training data. In this study, the Synthetic Minority Over-sampling Technique (SMOTE) was used because of its ability to produce realistic synthetic data through feature vector interpolation from the minority class [19]. SMOTE is widely used in the health sector because it does not simply duplicate data but creates variations that can help improve model generalization without overfitting [20], [21]. [22].

While SMOTE is applied to address class imbalance during training, the overall classification performance is also influenced by the representational quality of the input features. Most studies still rely solely on classical tabular-based features, which are often inadequate for capturing the complex, non-linear patterns associated with child growth trajectories. Conversely, DL models like RNN or CNNs can extract complex, non-linear features and uncover latent patterns embedded in data [23], [24], [25]. Unlike classical ML approaches that are handcrafted, CNN features are extracted automatically through a hierarchical learning process from neural networks [26]. Meanwhile, CNN models are more widely used as feature extractors in image data. In some CNN-based studies, images are only used as feature extractors, and ML classifiers such as SVM or RF are used. In 1D tabular data, CNN can be utilized, which has been proven effective in capturing local structures and sequence patterns [24], [25], [27]. CNN is used for feature extraction and can be combined with the final classification layer in one end-to-end network to improve overall performance. Therefore, combining classical interpretable features with deep features extracted from 1D-CNNs is a potential approach to improve the accuracy and generalization ability of the model. However, this fusion must be done strategically to maximize the information from each feature type. In this context, fusion methods become important to combine both strengths effectively.

To overcome the limitations of a single method, this study proposes a two-layer hybrid approach: intermediate fusion and late fusion. In the intermediate fusion stage, features extracted from 1D-CNN are combined with classical anthropometric features. The fusion is done through the attention function, specifically, Additive Attention, which is designed to focus weighting on the most relevant features in tabular data [28], [29]. Theoretically, the attention function allows the model to allocate higher weights to important features by considering correlation and contextual significance. Afterwards, in the late fusion stage, the classification results from several base classifiers are combined in an ensemble to produce a more robust final decision[14].

This two-level approach is inspired by Boulahia et al. [30], who compared and proved that intermediate and late fusion performs better than early fusion. In the health sector, the intermediate fusion strategy also positively affects modular data [31]. Building on these insights, this study hypothesizes that a dual-fusion architecture can substantially enhance model performance by leveraging deep and classical feature representations. To address the issue of class imbalance in health datasets, this study employs the SMOTE during the training phase. This approach ensures that the model adequately learns underrepresented classes, such as moderate and severe malnutrition. Building upon this balanced foundation, the study develops a classification model for predicting the nutritional status of toddlers using a hybrid dual-fusion approach. At the intermediate fusion stage, deep features extracted by 1D-CNN are combined with classical anthropometric features. These combined features are further refined using an additive attention mechanism, which assigns higher weights to the most informative attributes. This fusion enhances the model's capacity to represent complex nutritional patterns. At the late fusion stage, the classification outcomes from several base learners are integrated through an ensemble stacking strategy to ensure robustness and improved generalization. The combination of interpretable classical features and automatically learned deep features aims to maximize prediction accuracy. To evaluate the performance of the proposed model, key metrics such as accuracy, precision, recall, and F1-score are employed on a real-world dataset of toddlers.

2. Related Work

Several previous studies have examined the prediction of child malnutrition using individual ML algorithms. Most research on child nutritional status prediction still relies on individual ML algorithms. Studies by [32] and [33] reviewed the application of ML in the context of precision nutrition and general nutrition research, including the algorithms and metrics used. A specific predictive study was conducted by [34] in India using the AutoML approach and a combination of anthropometric indices such as stunting and wasting. At the same time, [35] evaluated the nutritional prediction of toddlers in Bangladesh and reported the best performance of the Random Forest algorithm. Other studies by [36] and [37] compared the performance of gradient boosting, logistic regression, and C4.5.

A study in Afghanistan [38] showed that Gradient Boosting (GB) excels in classifying three nutritional categories: undernutrition, stunting, and wasting. Several ensemble learning-based approaches have been used to improve prediction accuracy. For example, stacking was used in glaucoma classification by [39], clinical prediction of skin diseases by [17], detection of hypertension [40], and diabetes classification by [14]. Ensemble models have been shown to improve stability and accuracy by reducing variance and bias by combining multiple base models [41]. A recent paper by Jain et al. [42] further evaluated the efficiency of tabular ML and AutoML models in child malnutrition prediction in India, using a combination of HAZ, WHZ, and hybrid stunted-wasting (HAWH) indices. They concluded that AutoML and deep learning-based models such as TabNet provide high accuracy of up to 96.46% and AUC of 97.19%. However, they are limited to the automatic combination of classical models without architectural integration or deeper feature representation. A study by Duyar et al. [43] also showed that 1D-CNN outperformed other models in cardiovascular disease classification using tabular microbiota data, indicating the great potential of 1D CNN in the medical tabular domain.

In feature fusion, the common approaches used are concatenation and attention-based fusion. Concatenation techniques combine features from different sources into a single vector representation but often assume equal feature contributions. In contrast, attention-based approaches, such as additive attention, allow the model to give greater weight to contextual and relevant features, making it more adaptive to the complexity of tabular data. While existing studies have explored deep learning or ensemble approaches independently, the integration of 1D-CNN for deep feature extraction, attention-based intermediate fusion, and ensemble-based late fusion remains relatively underexplored in the context of nutritional status classification among toddlers. This study addresses this methodological gap by proposing a dual-fusion hybrid framework that aims to enhance classification accuracy through a synergistic combination of these techniques.

3. Proposed Method

This study proposes a classification model for the nutritional status of toddlers based on a two-level hybrid approach, namely intermediate fusion and late fusion. This model combines original features with deep feature representations from 1D CNN networks, which are then combined using additive attention and classified using an ensemble stacking model. The complete architecture is shown in figure 1.

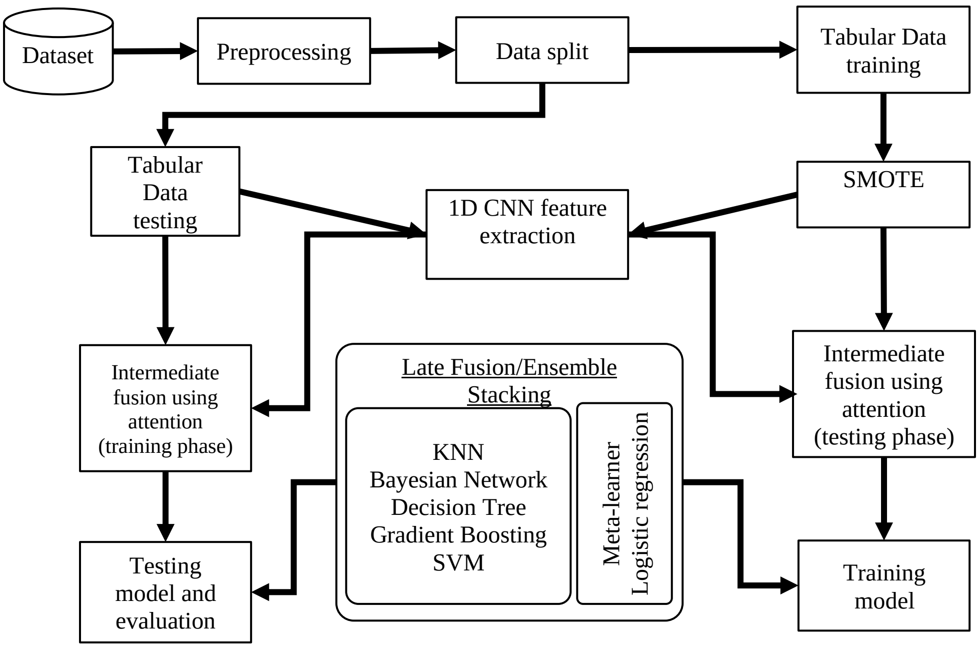


Figure 1. Proposed method

Figure 1 illustrates the overall pipeline of the proposed dual-fusion architecture. The process begins with data preprocessing, including cleaning, normalization, and stratified splitting. SMOTE is applied only to the training set to balance class distribution, ensuring that rare classes are adequately represented during learning. Afterwards, both training and testing data undergo a shared feature extraction and fusion process. Specifically, tabular features are passed through a 1D CNN to extract deep representations. These are then combined with classical anthropometric features using an additive attention mechanism. To clarify the modelling phases, the flowchart distinguishes intermediate fusion using attention (training phase) from intermediate fusion using attention (testing phase). This ensures transparency that fusion is applied identically during inference, but training involves SMOTE-balanced inputs.

In the final stage, the attention-weighted feature representations are classified using a stacking ensemble, including KNN, Bayesian Network, Decision Tree, Gradient Boosting, and SVM as base learners. Logistic Regression (LR) serves as the meta-learner.

3.1. Data Input and Preprocessing

The dataset used in this study comprises anthropometric and demographic characteristics of children under five years old, including weight, height, age in months, gender (categorical), and nutritional status as the class label. Several preprocessing steps were carried out to ensure the data's quality and reliability. First, the data cleaning process involved standard data frame operations to remove duplicate entries and records containing missing values in key columns. This step was essential to maintain the integrity of the dataset used for training and evaluation. Next, the categorical gender attribute was encoded into a binary format using one-hot encoding. To avoid issues of multicollinearity, only one of the resulting binary columns was retained—specifically, a column denoting gender as female (Gender_g), where a value of 1 represents a girl and 0 represents a boy.

Following encoding, all numerical features were normalized using standard z-score normalization to ensure that each feature contributes proportionately to the learning process. The normalization was performed using Equation (1):

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

x is the original value, μ is the mean of the feature, and σ is the standard deviation of the feature. Lastly, the dataset was split into 80% for training and 20% for testing using stratified sampling, which was conducted based on the nutritional status labels to preserve the original distribution of classes in both subsets. This stratification ensures that each class is proportionally represented, thereby supporting unbiased model evaluation.

3.2. Data Balancing with SMOTE

To address the class imbalance in the training data, this study applies the SMOTE. SMOTE generates synthetic samples for the minority class by interpolating between feature vectors of neighbouring minority instances in the feature space. We adopt the commonly used default setting of k_neighbors = 5, which defines the number of nearest neighbors used in the interpolation process[44]. This value has been widely used in literature for imbalanced classification tasks as a trade-off between oversampling aggressiveness and noise sensitivity. Given the moderate imbalance level in our dataset, this setting was deemed appropriate and effective for generating representative synthetic samples. SMOTE is applied exclusively to the training set after the data split to prevent data leakage, ensuring that the test data remains in its original, naturally imbalanced distribution.

3.3. 1D CNN Feature Extraction

To extract high-level representations from the tabular input $X \in \mathbb{R}^{n \times d2}$, a 1D-CNN is employed. CNN is effective in processing 1D structured data by capturing local patterns and sequence dependencies along the feature dimension [24], [25], [27]. Each row $x \in \mathbb{R}^{n \times d}$ of the input matrix is treated as a 1D sequence and reshaped to a tensor of shape $(d_2, 1)$ enabling the convolutional filters to slide over the feature dimension. This reshaping process allows 1D-CNN to learn spatial hierarchies of features by applying a series of convolution and pooling operations. The general form of the convolution operation applied to each input vector is defined as:

 ISSN 2723-6471

$$\mathrm{h}_i = f\{\sum_{j=0}^{k-1} w_j \cdot x_{i+j} + b\} \tag{2}$$

f is a non-linear activation (ReLU). w_j are the convolutional weights, and b is the bias term. The resulting output forms a deep feature map $D = [h_1, h_2, \ldots, h_m]$ is used as the deep-learned representation of the input instance; see Algorithm 1 for more details.

Algorithm 1: CNN Feature Extraction for Tabular Row Input

Input: Tabular data $X \in \mathbb{R}^{n\times d}$
Output: deep feature vector D
1: For each sample x in X:
2: Reshape $x \rightarrow (d, 1)$
3 Apply Conv1D → ReLU → BatchNorm
4: Apply MaxPooling
5: Apply second Conv1D → ReLU
6: Apply GlobalAveragePooling
7: Apply Dense(64) → ReLU

The detailed CNN configuration used in our method is presented in table 1.

Table 1. CNN 1D Architecture

Configuration	Value
Input	Vector with length equal to the number of original features
Conv1D layer 1	64 filters, kernel size = 3, activation = ReLU
BatchNormalization	Applied after the first convolution layer to stabilize learning
MaxPooling1D	Pool size = 2
Conv1D layer 2	128 filter, kernel size = 3, activation = ReLU
GlobalAveragePooling1D	Converts feature maps to a one-dimensional vector.
Dense	64 unit, activation ReLU
Optimizer	Adam, learning rate: 0.001
Epoch	50 with early stopping.
Batch size	32

The extracted deep features D are subsequently integrated with classical features in the intermediate fusion stage described in the following subsection.

3.4. Intermediate Fusion with Additive Attention

Following feature extraction, the classical tabular features $C \in \mathbb{R}^{d1}$ are concatenated with the deep CNN feature vector D to form a joint representation $F = [C||D]$. Instead of relying solely on concatenation, an additive attention mechanism is employed to selectively emphasize the most informative features from the combined representation.

Mathematically, the additive attention mechanism, as applied in recent attention-based deep fusion frameworks, computes a context vector c from the combined feature set $F = [F_1, F_2, \ldots, F_n]$. Each F_i is transformed via a non-linear projection and scored accordingly [29].

$$e_i = v^{\mathrm{T}} \tanh(W \cdot F_i + b) \tag{3}$$

The scores e_i are normalized via softmax:

$$\alpha_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)} \tag{4}$$

The context vector c is then computed as a weighted sum:

$$c = \sum_i \alpha_i \cdot F_i \tag{5}$$

This mechanism allows the model to focus on salient dimensions across classical and learned features, resulting in a more informative and discriminative representation. The complete fusion process is summarized in Algorithm 2.

Algorithm 2: Additive Attention-Based Intermediate Fusion

Input:
- Classical features $C \in \mathbb{R}^{d1}$
- Deep feature vector $D \in \mathbb{R}^{n \times d2}$

Output:
- Context vector $c \in \mathbb{R}^{d}$

1: Concatenate features: $F = [C||D]$
2: for each feature vector F_i in F do
3: $e_i = v^{\mathrm{T}} tanh(W \cdot F_i + b)$
4: end for
5: Compute attention weights:
6: $\alpha_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)}$
7: Compute context vector:
8: $c = \sum \alpha_i \cdot F_i$
9: return c

To further support reproducibility, the implementation of the attention mechanism is detailed in table 2.

Table 2. Additive Attention Architecture

Configuration	**Value**
Input	Concatenated CNN deep features and original features
Dense layer	128 units, activation = tanh
Dense layer	1 unit, linear activation to compute attention scores
Softmax	Converts attention scores into weight distribution
Output	Focused feature representation based on learned attention weights

Table 2 delineates the architectural configuration of the hybrid Attention-CNN model. The attention mechanism emphasizes salient feature representations extracted by the convolutional layers, which are subsequently processed through a dual-layered fully connected neural structure to enhance feature abstraction. The resulting feature vector is then propagated to a softmax activation function, which computes a normalized probability distribution over the output classes, thereby facilitating the final classification decision.

3.5.Ensemble Classification/Late Fusion

The proposed model employs a stacking ensemble strategy to finalize classification, where multiple base classifiers are combined through a meta-learner to enhance predictive performance and generalization. The base learners used in this study include KNN, Bayesian Network, DT, GB, and SVM. These classifiers are chosen for their complementary learning characteristics. Their configurations are listed in table 3.

Table 3. Classifier Model Architecture

Configuration	**Value**
KNN	n_neighbors = 5
Bayesian Network	default

Configuration	Value
DT	max_depth = 5
GB	n_estimators = 100, learning_rate = 0.1
SVM	kernel = 'rbf', C = 1.0
LR (meta learner)	Solver: liblinear; Regulation: L2 (default)

During the training phase, a 5-fold cross-validation stacking procedure is adopted. The base learners are trained using four folds of the training set in each fold. Their predictions on the remaining (held-out) fold are collected as out-of-fold predictions, forming the input features for the meta-learner. This process ensures that the meta-learner is trained on data that was never seen by the base learners during their individual training, thus avoiding information leakage. We adopt a stacked generalization strategy based on learned aggregation, in which a Logistic Regression model is trained as a meta-learner on the out-of-fold predictions. This approach is chosen over soft-voting or hard-voting strategies because it allows the meta-learner to model interactions and relative confidence levels among base models, leading to more effective combination of their strengths and improved generalization. During inference, all base learners are retrained on the full training set. The predictions generated by these retrained models on the test set are passed to the trained meta-learner, which then outputs the final classification result. This late fusion architecture leverages the diversity of base models and allows the meta-learner to adaptively weigh their outputs, leading to more stable and accurate predictions.

3.6.Model Testing and Evaluation

In the testing phase, the testing data is processed using the same pipeline: features are extracted using 1D-CNN, merged via attention, and classified using the trained stacking model. Performance evaluation uses metrics, namely accuracy, precision, recall, F1-score, and confusion matrix. Table 4 presents a concise input–process–output overview of the proposed pipeline to summarize the proposed model workflow.

Table 4. Model Pipeline Overview (Input–Process–Output Summary)

Stage	Component	Description
Input	Anthropometric & Demographic Features	Weight, height, age (in months), birth weight, birth height, gender (binary)
	Class Label	Nutritional status (e.g., stunted / not stunted)
Process	Preprocessing	Missing value removal, duplicate removal, normalization (StandardScaler), one-hot encoding
	Data Balancing (Training only)	SMOTE (k = 5) applied post-split on the minority class
	Feature Extraction	1D CNN with two convolution layers and pooling (table 1 configuration)
	Intermediate Fusion	Additive attention mechanism on concatenated deep and classical features
	Late Fusion/ Ensemble Classification	Ensemble stacking: KNN, Bayesian Network, DT, GB, SVM, and meta learner: Logistic Regression
Output	Prediction	Binary classification: predicted nutritional status label
	Evaluation	Accuracy, Precision, Recall, F1-Score, Confusion Matrix

4. Results and Discussion

4.1. Proposed Method Results

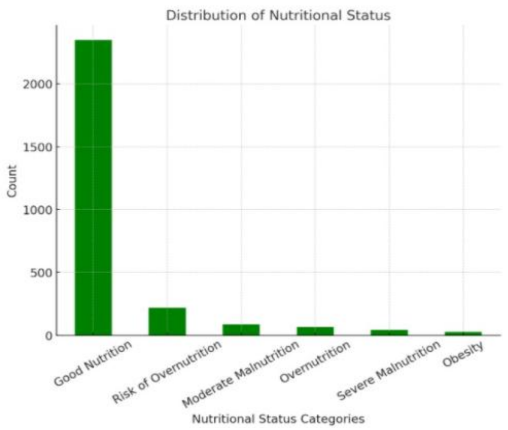
The dataset for this study was obtained from *posyandu* (integrated health service posts) in Semarang City, Central Java, Indonesia, consisting of 2,789 children under five. The data collection was carefully considered to ensure a representative sample that covers various child health indicators, including birth weight, birth height, age, weight, height, gender, and nutritional status. This comprehensive dataset provided a solid foundation for the subsequent analysis. Before the analysis, extensive data preprocessing was conducted to ensure the dataset was clean, consistent, and ready for modelling. The preprocessing steps involved handling missing values, deleting duplicates, and normalization. The data after cleaning and handling missing values is shown in table 5.

The categorical feature, gender, was then converted into a binary format using one-hot encoding. This step was crucial to make the data compatible with ML algorithms, as most models require numerical input. The encoding process resulted in new binary columns representing each category in the gender feature, which were then integrated back into the dataset. The data conversion results in one-hot encoding, as shown in table 6.

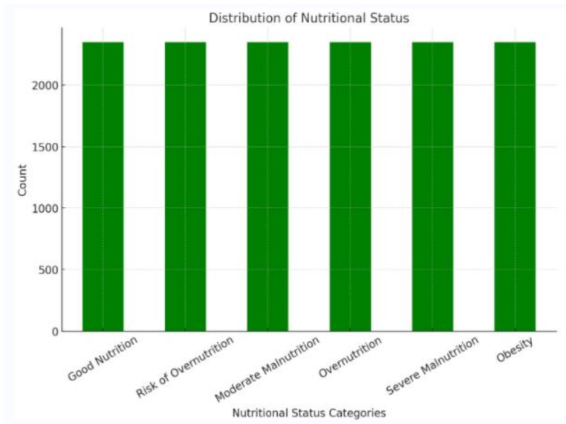
Table 5. Sample Data Before and After Gender Encoding

Birth Weight (kg)	Birth Height (cm)	Age (months)	Weight (kg)	Height (cm)	Nutritional Status	Gender_g	Gender_g (Encoded)
3.0	45.0	1	4.9	55.0	Good	girl	1
3.2	49.0	1	4.5	55.6	Good	girl	1
3.0	49.0	2	4.6	55.7	Good	girl	1
3.2	49.0	2	5.8	62.1	Good	boy	0
2.8	48.0	2	4.2	55.5	Good	boy	0

Data balancing with oversampling helps to make the data more reliable by avoiding overfitting or underfitting the model. This technique ensures that the data represents the overall population, reducing the risk of biased or skewed results. The initial dataset consisted of 2,789 children under 5 years of age records, which included the following values: Good Nutrition, Moderate Malnutrition, Obesity, Overnutrition, Risk of Overnutrition, and Severe Malnutrition. After applying the oversampling method, the data was balanced to match the largest class, as shown in figure 2 (a) and (b).



(a)



(b)

Figure 2. Original (a) and balanced data (b)

After the SMOTE oversampling process, the data distribution in each nutritional status category becomes balanced, as shown in figure 2 (b). This is important to ensure that the classification model is not biased towards the majority class and has a fair ability to recognize all categories, including minority classes such as Severe Malnutrition and Obesity. The next step is training and testing the model based on the proposed pipeline. The training model uses balanced data, while the evaluation is based on unmodified test data. The two-level hybrid model, which consists of intermediate and late fusion, is tested against several evaluation metrics, including accuracy, precision, recall, F1-score, and 5-fold cross-validation. This evaluation is carried out to ensure the model's prediction accuracy and the balance of sensitivity and specificity between classes. The results are shown in table 6 and the confidence matrix in figure 2 (b).

Table 6. Evaluation Results of the Proposed Method

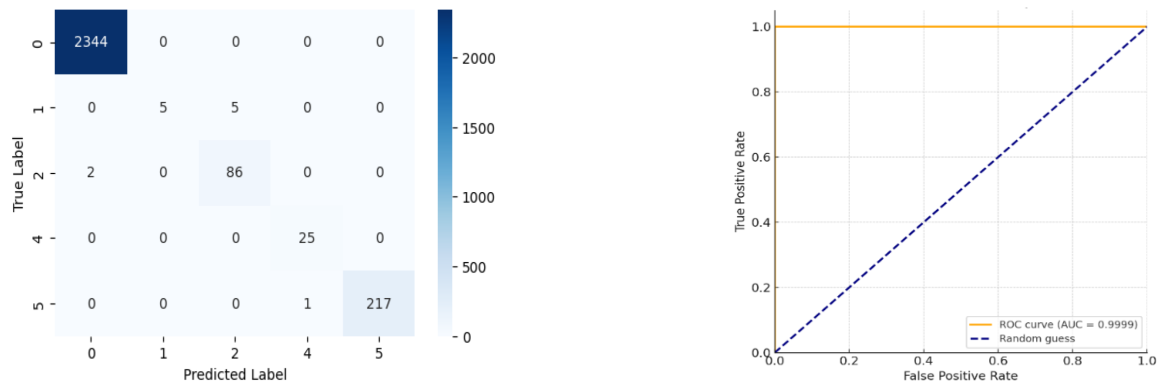
Fold	Acc	Pre	Rec	F1	Spe	AUC
1	99.44	99.43	99.44	99.40	1.0	99.99
2	99.81	99.82	99.81	99.78	1.0	99.99

3	99.81	99.82	99.81	99.78	1.0	99.99
5	99.62	99.67	99.63	99.60	1.0	99.99
5	99.81	99.82	99.81	99.78	1.0	99.99
Average	99.70	99.71	99.70	99.67	1.0	99.99
Std dev	0.17	0.17	0.16	0.17	0	0
95% CI	[99.57–99.83]	[99.54–99.88]	[99.57–99.83]	[99.51–99.83]	-	-

The evaluation results in table 6 show that the proposed model consistently achieved high performance across all 5-fold cross-validation runs. The average accuracy reached 99.70%, with precision, recall, and F1-score values of 99.71%, 99.70%, and 99.67%, respectively. The standard deviations for all metrics are below 0.2, indicating minimal variability and suggesting stable model behaviour across folds. Furthermore, the 95% confidence intervals for each metric affirm the robustness and statistical reliability of the results. In addition, the AUC value of 99.99% across all folds confirms the model's excellent discriminatory capability, while the perfect specificity (1.0) highlights its ability to avoid false positives—critical for avoiding misdiagnosis in nutritional assessments.

However, we acknowledge that such consistently high-performance metrics with low variance may raise concerns regarding potential overfitting or data leakage. To mitigate this, several precautions were taken. First, the SMOTE technique was applied strictly only to the training data after the train-test split, ensuring that no synthetic samples contaminated the test set. Second, stratified 5-fold cross-validation was employed to preserve class distribution while maintaining a strict separation between training and validation folds. Third, the test data used for evaluation remained in its original, imbalanced form, thereby reflecting real-world class proportions and minimizing the risk of inflated performance due to artificial balance.

To further validate the model's generalisation capability, we also conducted an independent hold-out test using a non-overlapping subset that was never exposed to the model during cross-validation or training. The model demonstrated consistent performance on this external set, reinforcing the reliability of our proposed pipeline. Additionally, we rigorously examined the dataset for any duplicates or data leakage artefacts during preprocessing. These preventive steps ensure that the model's performance is not a byproduct of overfitting but a reflection of its effective integration of deep features, attention-based fusion, and ensemble classification. Figure 3(a) presents the cumulative confusion matrix, while figure 3 (b) shows the ROC-AUC curve from the 5-fold cross-validation results. Most predictions align along the diagonal, indicating strong per-class accuracy. Although minor inter-class misclassifications are present—such as between class 1 (good nutrition) and class 2 (risk of overnutrition)—these instances are relatively infrequent and do not significantly degrade overall performance. This reinforces the hypothesis that the combination of SMOTE and dual-fusion architecture yields robust and informative feature representations capable of handling nuanced variations in class boundaries.



(a) (b)

Figure 3. The cumulative confusion matrix (a) ROC-AUC curve (b) obtained from 5-fold cross-validation

4.2. Ablation Studies

An ablation study was conducted on six model variants (M1–M6) to identify the contribution of each component in the proposed architecture, as presented in table 8. In addition to performance metrics, table 7 includes a new Structure

 ISSN 2723-6471

column that outlines the specific architectural flow for each variant, helping clarify the changes across configurations. In this experiment, Gradient Boosting was selected as the best baseline single classifier, as it previously showed superior performance compared to other models such as KNN, Bayesian Network, and SVM.

Table 7. Ablation study results on different model configurations

Model					Acc	Pre	Rec	F1	Spe	AUC
Variant	**Structure**	**SMOTE**	**CNN 1D**	**Ensemble**						
M1	Classical features → Gradient Boosting Classifier	✗	✗	✗	90.84	90.86	90.84	90.45	0.90	90.76
M2	Classical features → SMOTE → Gradient Boosting	✓	✗	✗	91.40	90.86	91.40	90.76	0.91	90.76
M3	Tabular input → CNN → Gradient Boosting Classifier	✗	✓	✗	97.58	97.60	97.58	97.41	0.97	98.94
M4	Classical features → SMOTE → Ensemble/ Late Fusion Classifier	✓	✗	✓	91.24	91.48	91.24	91.28	0.92	90.99
M5	Classical + CNN features → SMOTE → Intermediate Fusion → Gradient Boosting Classifier	✓	✓	✗	98.51	98.40	98.51	98.09	0.98	99.22
M6 Proposed	Classical + CNN features → SMOTE → Intermediate Fusion → Ensemble/ Late Fusion Classifier	✓	✓	✓	**99.70**	**99.71**	**99.70**	**99.67**	**1.0**	**99.99**

The baseline model (M1), which only uses classical features and Gradient Boosting without balancing or deep learning features, produces an accuracy of 90.84%, with an F1-score of 90.45% and an AUC of 90.76%. The improvement is immediately visible in M2, where the addition of SMOTE provides performance improvements in all metrics, proving the effectiveness of SMOTE in overcoming class imbalance. The M3 variant, which only relies on deep features from a 1D CNN with attention without SMOTE, shows a significant jump in performance, with an accuracy of 97.58%, an F1-score of 97.41%, and an AUC of 98.94%. This shows deep feature representation greatly improves classification capacity, even without data balancing. However, the specificity of this model is still relatively lower compared to the configuration using SMOTE.

M4, a combination of SMOTE and an ensemble classifier, produces an F1-score of 91.28% and an AUC of 90.99% but still does not surpass the performance of M3. This shows that feature representation is more crucial than combining models without relevant deep features. Further performance improvements is observed in M5, where the combination of SMOTE data balancing technique with 1D-CNN and Attention feature extractor results in an F1-score of 98.09% and an AUC of 99.22%. M6, which represents the proposed method, integrates all components, including 1D-CNN, with additive attention and ensemble stacking. This configuration achieves the highest performance with an accuracy of 99.70%, F1-score of 99.67%, AUC of 99.99%, and a perfect specificity of 1.0.

Key findings from this study show that the largest contribution comes from integrating deep features through 1D CNN and attention, which can capture non-linear representations more effectively than classical features. In addition, using SMOTE plays a significant role in maintaining the balance of performance between classes, while the ensemble stacking strategy strengthens the stability and generalization of the overall model. The combination of the three provides the most optimal and robust results.

4.3. Comparison with Other Methods

Previous studies have evaluated using ML algorithms to classify children's nutritional status. For example, a study [6] employed a range of algorithms and reported that RF and LR achieved the best performance, each with an accuracy of 91.11% on the test data. In contrast, the SVM and KNN showed lower predictive accuracy, namely 77.61% and 71.10%, respectively. Similarly, in [7], Random Forest again demonstrated superior performance with an AUC of 0.756

ISSN 2723-6471

and an accuracy of 67.6%. In contrast, Generalized Linear Model (GLM) and Neural Network (NN) underperformed with accuracies of 37.1% and 64.6%, and AUC values of 0.697, respectively. A more recent study [42] from India utilized TabNet and reported high accuracy and AUC values of 96.46% and 97.19%, outperforming classical models such as LR, RF, XGBoost, and ANN. Table 8 summarizes the best-performing algorithms reported in prior research.

Table 8. Comparison with Other Methods

Ref	Best algorithm	Acc	AUC
[6]	RF/LR	91.11	-
[7]	RF	67.60	75.60
[42]	TabNet	96.46	97.19
Ours	Dual fusion+ SMOTE	99.71	99.99

While Random Forest emerges as a consistently strong baseline across various studies, and TabNet shows promising deep learning-based results, the proposed model in this study achieves a markedly higher performance, with an accuracy of 99.71% and an AUC of 99.99%. This significant improvement can be attributed to several key innovations in our method. First, the integration of 1D-CNN allows the model to capture complex, non-linear relationships that classical models may overlook, particularly in tabular health datasets. Second, additive attention enhances the model's ability to weigh feature importance dynamically, offering a more nuanced representation than simple feature concatenation. Third, the ensemble stacking strategy adds robustness by combining the strengths of diverse classifiers, mitigating the individual weaknesses of base models. It is also important to note that, unlike previous studies that relied solely on conventional tabular features or AutoML pipelines, our approach explicitly addresses class imbalance using SMOTE, ensuring that rare cases such as severe malnutrition are not underrepresented during training. Additionally, the model was evaluated on unaltered test data, preserving real-world class distributions and avoiding artificial performance inflation.

5. Conclusion

This study introduced a dual-fusion hybrid approach that integrates deep learning, attention-based intermediate fusion, and ensemble classification for predicting the nutritional status of children under five. Integrating 1D-CNN and additive attention effectively enhances feature representation beyond classical anthropometric data. SMOTE addressed the class imbalance, while ensemble stacking enhanced model robustness and generalization. The proposed model outperforms traditional ML methods, achieving an average accuracy of 99.70% and an AUC of 99.99%, surpassing several baseline and state-of-the-art models.

This framework shows strong potential for real-world deployment in community health settings, where early and accurate nutritional screening can support targeted interventions. However, the model's performance is evaluated on a regional dataset, which may limit generalizability across broader populations with different sociodemographic profiles or health conditions. For future research, extending the dataset to include national-level samples, integrating temporal growth trends, and exploring interpretability techniques such as SHAP or LIME could improve model reliability and explainability in public health decision-making. Real-time or mobile-based implementation is also a promising direction to enable wider accessibility and usability in low-resource environments.

6. Declarations

6.1. Author Contributions

Conceptualization: W.H., K.D.H.; Methodology: W.H., I.S.; Software: W.H.; Validation: K.D.H., C.A.; Formal Analysis: W.H.; Investigation: W.H.; Resources: K.D.H., I.S.; Data Curation: W.H.; Writing – Original Draft Preparation: W.H.; Writing – Review and Editing: K.D.H., I.S., C.A.; Visualization: W.H.; 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

The authors would like to acknowledge financial support from the Directorate of Research, Technology, and Community Service, Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology of Indonesia, which was provided through a grant with reference number 108/E5/PG.02.00.PL/2024.

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] 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," *Pediatr. Heal. Med. Ther.*, vol. 10, no. 10. pp. 115–130, Oct. 2019, doi: 10.2147/PHMT.S222198.

[2] 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. 16-38,Sep. 2023, doi: 10.3390/children10101638.

[3] M.A.L. Suratri, G. Putro, B. Rachmat, Nurhayati, Ristrini, N.E. Pracoyo, A. Yulianto, A. Suryatma, M. Samsudin, Raharni, "Risk Factors for Stunting among Children under Five Years in the Province of East Nusa Tenggara (NTT), Indonesia," *Int. J. Environ*. Res. Public Health, vol. 20, no. 2, pp. 16-38,Jan. 2023, doi: 10.3390/ijerph20021640.

[4] W. Hadikurniawati, K. D. Hartomo, and I. Sembiring, "Spatial Clustering of Child Malnutrition in Central Java: A Comparative Analysis Using K-Means and DBSCAN," *in 2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)*, vol. 2023, no. 11. pp. 242–247, Nov. 2023, doi: 10.1109/ICMERALDA60125.2023.10458202.

[5] G. Rezaeizadeh et al., "Maternal education and its influence on child growth and nutritional status during the first two years of life: a systematic review and meta-analysis," *eClinicalMedicine*, vol. 71, no. 1, pp. 10-25, May 2024, doi: 10.1016/j.eclinm.2024.102574.

[6] S. Kar, S. Pratihar, S. Nayak, S. Bal, G. H L, and R. V, "Prediction of Child Malnutrition using Machine Learning," *in 2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)*, vol. 2021, no. 12. pp. 01–04, Dec. 2021, doi: 10.1109/IEMECON53809.2021.9689083.

[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 Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 291, Oct. 2021, doi: 10.1186/s12911-021-01652-1.

[8] H. Alawneh and A. Hasasneh, "Survival Prediction of Children after Bone Marrow Transplant Using Machine Learning Algorithms," *Int. Arab J. Inf. Technol.*, vol. 21, no. 3, pp. 394-407, 2024, doi: 10.34028/iajit/21/3/4.

[9] Q. Liu, M. Zhang, Y. He, L. Zhang, J. Zou, Y. Yan, Y. Guo, "Predicting the Risk of Incident Type 2 Diabetes Mellitus in Chinese Elderly Using Machine Learning Techniques," *J. Pers. Med.*, vol. 12, no. 6, pp. 905, May 2022, doi: 10.3390/jpm12060905.

[10] O.N. Chilyabanyama, R. Chilengi, M. Simuyandi, C.C. Chisenga, M. Chirwa, K. Hamusonde, R.K. Saroj, N.T. Iqbal, I. Ngaruye, S. Bosomprah, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, pp. 1082, Jul. 2022, doi: 10.3390/children9071082.

[11] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, "Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia," *Public Health Nutr.*, vol. 25, no 2, pp. 269 - 280, Feb. 2022, doi: 10.1017/S1368980021004262.

[12] R. Gustriansyah, N. Suhandi, S. Puspasari, and A. Sanmorino, "Machine Learning Method to Predict the Toddlers' Nutritional Status," *J. INFOTEL*, vol. 16, no. 1, pp. 32-43, Jan. 2024, doi: 10.20895/infotel.v15i4.988.

[13] T. Hariguna, "Health and Socio-Demographic Risk Factors of Childhood Stunting: Assessing the Role of Factor Interactions Through the Development of an AI Predictive Model," *J. Appl. Data Sci.*, vol. 5, no. 4, pp. 2175–2186, Dec. 2024, doi: 10.47738/jads.v5i4.612.

[14] 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," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 23–38, May 2024, doi: 10.62411/faith.2024-11.

[15] J. P. Ntayagabiri, Y. Bentaleb, J. Ndikumagenge, and H. El Makhtoum, "OMIC: A Bagging-Based Ensemble Learning Framework for Large-Scale IoT Intrusion Detection," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 401–416, Feb. 2025, doi: 10.62411/faith.3048-3719-63.

[16] A. Çetin and S. Öztürk, "Comprehensive Exploration of Ensemble Machine Learning Techniques for IoT Cybersecurity Across Multi-Class and Binary Classification Tasks," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, pp. 371–384, Feb. 2025, doi: 10.62411/faith.3048-3719-51.

[17] O. Jaiyeoba, E. Ogbuju, O. T. Yomi, and F. Oladipo, "Development of a model to classify skin diseases using stacking ensemble machine learning techniques," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 22–38, May 2024, doi: 10.62411/jcta.10488.

[18] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, no. August, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[19] F.O. Aghware, A.A. Ojugo, W. Adigwe, C.C. Odiakaose, E.O. Ojei, N.C. Ashioba, M.D. Okpor, V.O. Geteloma, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 407–420, Mar. 2024, doi: 10.62411/jcta.10323.

[20] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, "Enhancing SMOTE for imbalanced data with abnormal minority instances," *Mach. Learn. with Appl.*, vol. 18, no. 12, pp. 100597, Dec. 2024, doi: 10.1016/j.mlwa.2024.100597.

[21] D. R. I. M. Setiadi, H. M. M. Islam, G. A. Trisnapradika, and W. Herowati, "Analyzing Preprocessing Impact on Machine Learning Classifiers for Cryotherapy and Immunotherapy Dataset," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 39–50, Jun. 2024, doi: 10.62411/faith.2024-2.

[22] T. Sugihartono, "Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 667–682, Jan. 2024, doi: 10.47738/jads.v6i1.494.

[23] D.R.I.M. Setiadi, A.A. Ojugo, O. Pribadi, E. Kartikadarma, B.H. Setyoko, S. Widiono, R. Robet, T.C. Aghaunor, E.V. Ugbotu, "Integrating Hybrid Statistical and Unsupervised LSTM-Guided Feature Extraction for Breast Cancer Detection," *J. Comput. Theor. Appl.*, vol. 2, no. 4, pp. 536–552, May 2025, doi: 10.62411/jcta.12698.

[24] D. Sunaryono, R. Sarno, J. Siswantoro, D. Purwitasari, S.I. Sabilla, R.I. Susilo, N.R. Akbar, "Hybrid One-Dimensional CNN and DNN Model for Classification Epileptic Seizure," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 6, pp. 492–502, Dec. 2022, doi: 10.22266/ijies2022.1231.44.

[25] K. Anam, "Early Stopping on CNN-LSTM Development to Improve Classification Performance," *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1175–1188, Sep. 2024, doi: 10.47738/jads.v5i3.312.

[26] T. Sutojo, D. R. Ignatius, M. Setiadi, S. Rustad, W. Herowati, and M. Akrom, "Hybrid Quantum-Deep Learning Approach: Optimizing Land Cover Classification with GMM Outlier and Fusion Key Feature Selection," *Int. J. Intell. Eng. Syst.*, vol. 18, no. 1, pp. 638–648, Feb. 2025, doi: 10.22266/ijies2025.0229.45.

[27] C. Asuai, P. O. Ezzeh, H. Honi, A. A. Joseph-Brown, I. A. Merit, and I. Debekeme, "Enhancing DDoS Detection via 3ConFA Feature Fusion and 1D Convolutional Neural Networks," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 145–162, 2025, doi: 10.62411/faith.3048-3719-105.

[28] A. F. Thielmann, A. Reuter, T. Kneib, D. Rügamer, and B. Säfken, "Interpretable Additive Tabular Transformer Networks," *Trans. Mach. Learn. Res.*, vol. 2024, no. 5, pp. 1-19, 2024.

[29] N. El-Haryqy, A. Kharbouche, H. Ouamna, Z. Madini, and Y. Zouine, "Improved automatic modulation recognition using deep learning with additive attention," *Results Eng.*, vol. 26, no. 6, pp. 10-47, Jun. 2025, doi: 10.1016/j.rineng.2025.104783.

[30] S. Y. Boulahia, A. Amamra, M. R. Madi, and S. Daikh, "Early, intermediate and late fusion strategies for robust deep learning-based multimodal action recognition," *Mach. Vis. Appl.*, vol. 32, no. 6, pp. 121-134, Nov. 2021, doi: 10.1007/s00138-021-01249-8.

[31] A. N. Safriandono, D. R. I. M. Setiadi, A. Dahlan, F. Z. Rahmanti, I. S. Wibisono, and A. A. Ojugo, "Analyzing Quantum Feature Engineering and Balancing Strategies Effect on Liver Disease Classification," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 51–63, Jun. 2024, doi: 10.62411/faith.2024-12.

[32] D. Kirk, C. Catal, and B. Tekinerdogan, “Precision nutrition: A systematic literature review,” *Comput. Biol. Med.*, vol. 133, no. 6, pp. 10-43, Jun. 2021, doi: 10.1016/j.compbiomed.2021.104365.

[33] D. Kirk, E. Kok, M. Tufano, B. Tekinerdogan, E. J. M. Feskens, and G. Camps, “Machine Learning in Nutrition Research,” *Adv. Nutr.*, vol. 13, no. 6, pp. 2573–2589, Nov. 2022, doi: 10.1093/advances/nmac103.

[34] R. Joseph, V. Sawant, S. Shenai, M. Paryani, and G. Patil, “Machine Learning based Factors affecting Malnutrition and Anemia among children in India,” *in 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, vol. 2022, no. 5, pp. 1826–1833, May 2022, doi: 10.1109/ICICCS53718.2022.9788386.

[35] A. Talukder and B. Ahammed, “Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh,” *Nutrition*, vol. 78, no 10, pp. 11-26, Oct. 2020, doi: 10.1016/j.nut.2020.110861.

[36] Y. Baashar, G. Alkawsi, N. Ali, H. Alhussian, and H. T. Bahbouh, “Predicting student’s performance using machine learning methods: A systematic literature review,” *in 2021 International Conference on Computer & Information Sciences (ICCOINS)*, vol. 2021, no. 7, pp. 357–362 Jul. 2021. doi: 10.1109/ICCOINS49721.2021.9497185.

[37] A. Rosari, J. Julianto, A.D. Larasati, L.A. Pramesti, Triwiyanto, S. Lutfiyah, V. Abudlayev, “Developing a Nutritional Assessment Tool for Toddlers Using Anthropometry and IoT Technology,” *Int. J. Adv. Heal. Sci. Technol.*, vol. 4, no. 2, pp. 67–71, Apr. 2024, doi: 10.35882/ijahst.v4i2.319.

[38] A. Khan, S. Deshpande, and A. K. Tripathy, “Optimizing Nutrition using Machine Learning Algorithms-a Comparative Analysis,” *in 2019 International Conference on Nascent Technologies in Engineering (ICNTE)*, vol. 2019, no. 1, pp. 1–4, Jan. 2019. doi: 10.1109/ICNTE44896.2019.8946091.

[39] S. Pavithra, D. Jaladi, and K. Tamilarasi, “Optical imaging for diabetic retinopathy diagnosis and detection using ensemble models,” *Photodiagnosis Photodyn. Ther*., vol. 48, no. 8, pp. 10-425, Aug. 2024, doi: 10.1016/j.pdpdt.2024.104259.

[40] C.C. Odiakaose, F.O. Aghware, M.D. Okpor, A.O. Eboka, A.P. Binitie, A.A. Ojugo, D.R.I.M. Setiadi, A.E. Ibor, R.E. Ako, V.O. Geteloma, E.V. Ugbotu, T.C. Aghaunor, “Hypertension Detection via Tree-Based Stack Ensemble with SMOTE-Tomek Data Balance and XGBoost Meta-Learner,” *J. Futur. Artif. Intell. Technol*., vol. 1, no. 3, pp. 269–283, Dec. 2024, doi: 10.62411/faith.3048-3719-43.

[41] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, “HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System,” *IEEE Access*, vol. 8, no. 7, pp. 13-30–133050, 2020, doi: 10.1109/ACCESS.2020.3010511.

[42] S. Jain, T. Khanam, A. J. Abedi, and A. A. Khan, “Efficient Machine Learning for Malnutrition Prediction among under-five children in India,” *in 2022 IEEE Delhi Section Conference (DELCON)*, vol. 2022, no. 4, pp. 1–10, 2022, doi: 10.1109/DELCON54057.2022.9753080.

[43] C. Duyar, S. O. Senica, and H. Kalkan, “Detection of cardiovascular disease using explainable artificial intelligence and gut microbiota data,” *Intell. Med.*, vol. 10, no. April, pp. 10-18,2024, doi: 10.1016/j.ibmed.2024.100180.

[44] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning from Imbalanced Data Sets. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-98074-4.