IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 31, No. 7, July, 2026, pp. 2177-2188
Journal homepage: http://iieta.org/journals/isi

Explainable Multimodal Learning for Pediatric Vital Sign Anomaly Detection Using Physiological and Motion Signals

Adlane Djekhar[1], Naçima Mellal[2*], Imad Benacer[2]

[1] Research Laboratory of Artificial Intelligence and Autonomous Objects, University of Oum El Bouaghi, Oum El Bouaghi 04000, Algeria
[2] Department of Networks and Telecommunication, Institute of Technology, Research Laboratory of Artificial Intelligence and Autonomous Objects, University of Oum El Bouaghi, Oum El Bouaghi 04000, Algeria

Corresponding Author Email: nassima.mellal@univ-oeb.dz

https://doi.org/10.18280/isi.310705

Received: 19 April 2026
Revised: 10 July 2026
Accepted: 21 July 2026
Available online: 31 July 2026

Keywords:
pediatric monitoring, vital sign anomaly detection, explainable artificial intelligence, multimodal learning, physiological signals, XGBoost classifier, wearable healthcare

ABSTRACT

Pediatric physiological monitoring in non-clinical environments remains challenging due to limited integration of heterogeneous physiological signals and insufficient interpretability of automated decision-making systems. This study proposes an explainable multimodal learning framework for pediatric vital sign anomaly detection by integrating photoplethysmography (PPG), heart rate, oxygen saturation (SpO_2), body temperature, and inertial measurement unit (IMU) signals. Physiological, morphological, and motion-related features were extracted from 10-second signal windows and used to train an XGBoost-based classifier with cost-sensitive optimization to address class imbalance. The model was evaluated using stratified five-fold cross-validation and achieved a test accuracy of 93% and a macro F1-score of 0.90 across seven physiological categories. Compared with single-modality models, the proposed multimodal framework improved classification performance by 5% over PPG-only inputs and by more than 20% over temperature- and IMU-based models. Shapley Additive Explanations (SHAP)-based analysis demonstrated consistent feature contribution patterns, with feature importance rankings showing strong agreement across validation folds (Jaccard index = 0.76; Spearman correlation = 0.96). The proposed framework provides an interpretable and computationally efficient approach for pediatric physiological anomaly screening and may support future IoT-based monitoring applications following prospective clinical validation.

1. INTRODUCTION

The detection of abnormalities based on vital signs is a key challenge in continuous medical monitoring. This issue is of particular importance for children in group settings, such as nurseries or daycare centers, where constant individual monitoring remains limited and insufficient. In this context, automatic and early detection of physiological abnormalities—such as unusual variations in heart rate, oxygen saturation (SpO_2), or motor activity—can contribute to rapid intervention and the reduction of potential complications. Numerous studies have explored this field, driven by the rapid evolution of wearable sensors and IoT infrastructure, which now enable the real-time acquisition of multiple physiological signals [1-3].

Artificial intelligence (AI) approaches, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), have been widely used in recent years to model temporal or morphological variations in biomedical signals. However, most of these models treat them independently. They do not fully exploit the relationships between multiple physiological parameters measured simultaneously, which limits their ability to identify more comprehensive and multimodal health anomalies. Recent studies highlight the need to move from monomodal analysis to multimodal analysis and data fusion for more accurate and earlier diagnosis [4].

In addition, many existing approaches achieve strong predictive performance but provide limited insight into how decisions are made. Methods such as Shapley Additive Explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM) are available to improve interpretability [5], yet they are not systematically integrated into current models. This lack of transparency can slow down clinical adoption. The issue becomes even more critical in pediatric contexts, where decisions must be clearly explained to both healthcare professionals and caregivers given the sensitivity and responsibility associated with monitoring young patients.

In this context, the prioritized integration of data from multimodal sensors, combined with the use of an interpretable AI model, appears to be a promising approach for detecting abnormalities in children in a group setting. By combining complementary signals such as heart rate, oxygen saturation, and body temperature with photoplethysmography (PPG) and motion data from an accelerometer and a gyroscope, it becomes possible to improve robustness against noise and obtain a more comprehensive representation of the child's physiological state.

However, despite the growing body of work in this area, existing approaches present three recurring limitations. First, most studies focus on isolated physiological modalities and do not exploit the complementary information available across simultaneously acquired signals, limiting robustness in the presence of inter-individual variability. Second, the majority of AI-based monitoring systems operate as black-box models, with limited or no integration of explainability mechanisms, which restricts their clinical adoption in sensitive pediatric contexts. Third, most validated frameworks target adult or hospital populations, leaving the specific physiological and behavioral characteristics of young children in non-clinical daycare settings largely unaddressed.

This study examines the use of the XGBoost model for the detection of physiological anomalies from multimodal signals by integrating an interpretability mechanism based on the SHAP method, in order to ensure the transparency of the model's decisions, precisely quantify the contribution of each physiological variable to predictions, and strengthen clinical confidence in a sensitive pediatric context.

The main contributions of this work can be summarized as follows:

(1). A multimodal physiological anomaly detection framework that simultaneously integrates multiple vital signs (heart rate, SpO_2, temperature, etc.), enabling us to overcome the limitations of single-signal-based approaches and improve robustness against physiological variability.

(2). The design and optimization of an interpretable XGBoost model, ensuring transparency in the model's decision-making.

(3). The design of a pediatric-specific monitoring framework that targets health-state assessment in childcare environments, a context that remains less explored than adult healthcare monitoring.

(4). Beyond the use of XGBoost and SHAP, an important aspect of this work is the development of a lightweight and interpretable monitoring framework intended for future deployment in resource-constrained pediatric IoT environments. The choice of XGBoost was motivated not only by predictive performance but also by its relatively low computational requirements compared with many deep learning approaches. While deployment on edge hardware was not evaluated in the present study, the proposed architecture was developed with this objective in mind and will be investigated in future work.

In the remainder of this study, we review previous studies in the second section, discuss the methodology—including the dataset and the algorithm used—in the third section, present the results and discussions in the fourth section, and conclude with a summary and directions for future work.

2. RELATED WORK

Research into AI algorithms for detecting vital sign abnormalities has become a critical field for the continuous assessment of health status and the early detection of diseases [1, 2]. The field has made significant progress thanks to the widespread adoption of wearable sensor technologies and IoT-compatible systems, which enable the acquisition of real-time physiological data and the application of advanced machine learning models to analyze complex time-series signals [1]. However, despite these advances, existing approaches still present notable gaps in terms of multimodal integration, pediatric specificity, and decision transparency — three dimensions that the present work specifically addresses.

2.1 XGBoost for vital sign anomaly detection

Boosted decision tree models, such as XGBoost, have been widely used in anomaly detection and the prediction of critical events in connected health due to their ability to handle heterogeneous data and deliver high performance [6, 7].

In this context, Divyabharathi and Madhavan [8] proposed the Decision-Integrated Anomaly Detection and Emergency Management framework with eXplainable (DIADEM-X) models, which combines XGBoost and SHAP for accurate and interpretable detection, with high performance and low latency, making it suitable for edge computing. Another study also used XGBoost for sepsis prediction based on heterogeneous clinical data, demonstrating a notable improvement in performance area under the receiver operating characteristic curve (AUROC) [9]. Mayer et al. [10] presented the prediction of complications via circadian disturbances by integrating XGBoost and Synthetic Minority Over-sampling Technique (SMOTE), but obtained more moderate results.

While these studies confirm the effectiveness of XGBoost for health monitoring tasks, they target adult or mixed intensive care unit (ICU) populations and do not address the specific physiological characteristics of young children. Furthermore, none of them integrates multimodal physiological signals from wearable sensors in a non-clinical childcare setting. The present work addresses these gaps by applying XGBoost to a pediatric-specific feature space derived from simultaneously acquired PPG, heart rate, SpO_2, temperature, and inertial measurement unit (IMU) signals.

2.2 Pediatric multimodal monitoring

The use of data fusion to improve pediatric monitoring is not a recent development. For example, as reported by Shah [11], a system has been proposed that uses the PPG signal to automatically estimate respiratory rate and fuse multiple vital signs within a probabilistic framework, enabling early detection and assessment of infection severity. In a similar approach, methods for estimating respiration using pulse oximeters were developed, showing that data fusion significantly improves the performance of pediatric triage systems [12].

Furthermore, an intelligent multiparametric monitoring system was introduced in neonatal intensive care, based on unsupervised novelty detection techniques (Kernel Density Estimation (KDE), Kernel Principal Component Analysis (KPCA), One-Class Support Vector Machine (OCSVM)), to reduce false alarms. The results show a notable improvement in specificity while maintaining high sensitivity [13].

Although these works demonstrate the value of multimodal fusion for pediatric monitoring, they remain limited to clinical settings and lack any explainability mechanism for non-specialist caregivers.

2.3 Interpretable artificial intelligence in healthcare

Recent studies highlight the value of explainable AI (XAI) methods in improving the performance and interpretability of predictive models in pediatric healthcare.

Among these approaches, SHAP is particularly suitable for explaining feature contributions to individual predictions and overall model behavior, supporting its use with XGBoost for interpretable healthcare anomaly detection [14]. Previously, a multi-XAI approach (SHAP, LIME, ELI5, QLattice) was adopted to explain a random forest model, demonstrating that these techniques provide complementary insights into the factors influencing survival in respiratory diseases [15].

Furthermore, XGBoost and variants of SHAP have been applied to analyse the postoperative mortality risk and the early prediction of DIC [16, 17]. The results demonstrate high predictive performance and enable the identification of key biomarkers useful for clinical decision-making.

However, unlike the present work, none of these studies validate SHAP consistency quantitatively across cross-validation folds in a non-clinical monitoring context.

2.4 Internet of Things daycare and child monitoring

Smart monitoring of children's health and safety has also benefited from the growing integration of Internet of Things (IoT) technologies and machine learning in this field.

A multifunctional system was proposed that combines event detection, crying analysis, and facial expression analysis, although it is limited by environmental and computational constraints [18]. KiMS, a system based on wearable sensors and on-device processing to detect audio and physiological events, ensuring continuous monitoring via wireless communication [19].

Furthermore, an IoT solution was developed based on a connected wristband integrating physiological sensors and location tracking, offering reliable detection of critical events with a low margin of error [20]. In the study conducted by Pradeepa et al. [21], machine learning models are used for vital sign analysis, achieving high accuracy, particularly with support vector machine (SVM).

Despite advances in IoT- and AI-based health monitoring, several limitations remain in existing approaches. First, most studies lack joint integration of multimodal physiological signals for anomaly detection, instead focusing on isolated modalities, which limits their ability to capture complex interactions and reduces robustness in real-world conditions.

Moreover, a large number of studies rely on black-box AI models, with limited explainability mechanisms, and this diminishes clinical trust, complicates the interpretation of decisions, and limits their adoption in sensitive contexts involving children. Furthermore, a significant proportion of the research is validated on adult or mixed populations, with insufficient attention paid to pediatric cohorts and, more specifically, to daycare or nursery environments. These limitations—lack of multimodal integration, limited explainability, and insufficient pediatric specificity—define the research gaps that the present framework is designed to address. Table 1 presents a comparative overview of the reviewed studies and their results.

Table 1. Literature survey

Reference	Population / Context	Data and Sensors	Methods / Models	Key Results
Divyabharathi and Madhavan [8]	Patients monitored through IoT healthcare systems in hospitals, home-care environments, and mobile health units	Physiological signals collected from IoT sensors combined with patient historical records	XGBoost classifier, SHAP explainability, Emergency Decision Fusion Engine, Edge Computing architecture	Achieved 96.1% accuracy, F1-score of 94%, alert generation latency of 370 ms, and a false alarm rate of 3.5%
Lin et al. [9]	Adults (ICU – sepsis)	EHR, vital signs, laboratory data, interventions	XGBoost, temporal and count-based features	AUROC improvement of +5.4% (MIMIC-III) and +2.1% (PhysioNet 2019)
Mayer et al. [10]	Postoperative patients	Nocturnal vital signs	SMOTE + XGBoost	AUROC = 0.65; Precision = 0.75
Kumar et al. [15]	Pediatric respiratory diseases (survival prediction)	Hospital clinical data (ICU admission, potassium, creatinine, cyanosis, sodium levels)	Random Forest + SHAP, LIME, ELI5, QLattice	Accuracy = 96%
Kapsner et al. [16]	1,302 pediatric patients after congenital heart surgery	Postoperative laboratory data (serum creatinine, etc.)	Random Survival Forest + XGBoost + SHAP + SurvSHAP(t)	C-index = 0.85 (RSF) and 0.79 (XGBoost); maximum creatinine within 72 hours was highly predictive
Zhou et al. [17]	6,093 critically ill children (PICU) – DIC prediction	Biomarkers: D-dimer, INR, PT, TT, platelet count	XGBoost + SHAP	AUC = 0.908; Specificity = 85.9%;
Prathyanga et al. [18]	Children (daycare centers)	IoT sensors	Machine learning and IoT	Improved safety and supervision in daycare environments
Pradeepa et al. [21]	Students	IoT sensors (vital signs and behavioral data)	SVM, MLP, RF, DT	Maximum accuracy of 99.1% achieved with SVM
Our work	Children	23 features derived from PPG, heart rate, body temperature, and accelerometry (IMU)	XGBoost + SHAP	An accuracy of 93% and a macro F1-score of 0.90

Note: PPG: photoplethysmography; IMU: inertial measurement unit; IoT: Internet of Things; EHR: electronic health records; ICU: intensive care unit; PICU: pediatric intensive care unit; ASD: autism spectrum disorder; SHAP: Shapley Additive Explanations; SVM: support vector machine; MLP: multilayer perceptron; RF: random forest; DT: decision tree; SMOTE: Synthetic Minority Oversampling Technique; QLattice: Quantum Lattice; RSF: Random Survival Forest; DIC: disseminated intravascular coagulation; AUROC: area under the receiver operating characteristic curve; INR: international normalized ratio; PT: prothrombin time; TT: thrombin time.

3. METHODOLOGY

3.1 Dataset description

One of the main challenges in pediatric predictive analysis is the limited availability of high-resolution clinical datasets, mainly due to ethical and access constraints. In this context, a synthetic data-driven approach is adopted to enable the simulation of clinically relevant scenarios that are difficult to systematically capture in real hospital environments for this age group. This synthetic dataset is intended for pipeline development, methodological evaluation, and feature analysis, while clinical validation on real-world data remains necessary for generalization.

The generated dataset comprises 10,000 samples organized into 10-second time windows, simulating vital signs for children aged 1 to 5 years in accordance with the physiological ranges reported in the literature. Signals are sampled at two distinct resolutions: 1 Hz frequency for trend parameters, including heart rate, SpO_2, and temperature, and 100 Hz frequency for the PPG signal, to enable a detailed morphological analysis.

To ensure dataset realism, inter-child variability is introduced. Each simulated child has its own baseline values for heart rate, SpO_2, and temperature. These baseline values are chosen within normal pediatric ranges reported in medical literature (Table 2). This ensures that each child is different, as in real populations. Second, physiological rules are applied to model real relationships between vital signs. For example, fever increases heart rate, low oxygen levels often cause an increase in heart rate as a compensatory response, and low oxygen also reduces PPG signal strength. Movement affects the signal quality and makes the PPG signal noisier. These relationships are not fixed rules; small random variations are added to make the data more realistic. Third, time variation is added by generating short 10-second windows of signals. Within each window, the signals change over time instead of staying constant. Then, statistical features (such as mean, variation, and trend) are extracted, similar to those used in real IoT health monitoring systems.

The full dataset generation process can be seen as a simplified virtual monitoring system. For each sample, the system first selects a child with baseline characteristics, then assigns a physiological condition (such as fever, hypoxia, bradycardia, tachycardia, or motion artifact). After that, the system generates time-based signals that follow the physiological rules, adds realistic noise and artifacts, and finally extracts features used for machine learning training.

Table 2. Pediatric norms

Parameter	Pediatric Normal Range	Reference
Heart rate	80–130 bpm	Fleming et al. [22], WHO [23]
SpO_2	95–100%	WHO [23]
Bradycardia	<80 bpm	Fleming et al. [22], AHA [24]
Tachycardia	>130 bpm	AHA [24]
Mild hypoxia	90–94% SpO_2	WHO [23]
Severe hypoxia	<90% SpO_2	WHO [23]
Fever	>38 °C	AAP [25]
Temperature	approximately 37 °C	AAP [25]

Table 3. Distribution of classes in the dataset

Encoding	Class	Proportion
0	Normal	70%
1	Bradycardia	5%
2	Tachycardia	5%
3	Mild hypoxia	5%
4	Severe hypoxia	5%
5	Fever	5%
6	Motion / noise artifact	3%

The dataset is labeled into seven distinct classes (see Table 3), with the "Normal" class being the most prevalent (70%) to reflect the reality of standard monitoring.

Each record in the dataset is represented by 23 features, organized into several categories. The identifying information includes two variables: the child's ID and age. Heart rate parameters include six statistical descriptors: minimum, maximum, and mean values; standard deviation; slope; and interquartile range (IQR). Similarly, SpO_2 is characterized by six features, including minimum, maximum, mean, standard deviation, slope, and the number of desaturation events (drops greater than 3%). Body temperature is represented by three indicators: minimum, maximum, and mean values.

PPG signals are represented by four characteristics, including AC/DC components, the number of detected systolic peaks, and the signal quality index (SQI). Additionally, two supplementary characteristics describe the level of motion derived from the IMU and the overall signal quality. Finally, the dataset includes two output variables corresponding to the ID and the class label.

3.2 XGBoost algorithm

A crucial stage in creating an IoT-based physiological anomaly detection system is choosing the right AI model. In this context, the approach adopted relies on the use of the XGBoost classifier, whose suitability is justified by several technical considerations. On the one hand, this model is highly compatible with the SHAP interpretability method, thereby facilitating the analysis of variable contributions to predictions [26-28]. On the other hand, the data used in this study consist of tabular variables and derived features resulting from aggregation processes, which corresponds to a use case particularly well-suited to XGBoost [29, 30]. Furthermore, although the proposed multimodal fusion strategy relies on an early fusion of features, XGBoost can implicitly capture cross-modal interactions through its tree-partitioning mechanism, where decision nodes may jointly exploit features originating from different sensing modalities (e.g., IMU_motion_level and SpO_2_min). This capability allows the model to learn complex relationships between physiological and behavioral signals without requiring an explicit cross-modal fusion architecture. In addition, the target architecture incorporates an ESP32 and a Raspberry Pi, in which anomaly detection must be performed in real time while minimizing power consumption. This makes XGBoost a suitable choice for the proposed edge architecture because of its efficient prediction time and suitability for real-time healthcare applications [31].

The adopted XGBoost configuration represents a balance between nonlinear modeling capability, robustness to noise, and control of overfitting, while adhering to the computational constraints of a medical IoT environment. It is formulated for a multi-class classification task using a probabilistic objective function that outputs class probability distributions and is optimized using multi-class logarithmic loss. It is configured with a learning rate of 0.03 and 500 trees, with maximum depth limited to 5. Regularization is applied through L2 and L1 coefficients set to 2 and 0.5, respectively. Additionally, robustness is enhanced using a row subsampling ratio of 0.8 and a feature subsampling ratio of 0.8. To further address class imbalance and reduce the risk of overlooking clinically important anomalies, a cost-sensitive learning strategy is incorporated by assigning higher misclassification penalties to minority and critical health-condition classes during model training. Specifically, class weights are set inversely proportional to class frequency, such that $w_i = N / (K \times n_i)$, where, N is the total number of samples, K is the number of classes, and n_i is the number of samples in class c, ensuring that minority and clinically critical classes receive proportionally higher penalties during training.

To ensure a rigorous and statistically reliable evaluation, stratified five-fold cross-validation was employed, wherein the dataset is partitioned into five equally sized subsets while preserving the original class distribution. This approach is particularly suitable for anomaly detection tasks, which often involve class imbalance. At each iteration, four subsets are used for training and the remaining subset for validation, such that every observation serves exactly once as validation data. Prior to partitioning, the data are randomly shuffled to mitigate potential biases arising from their initial ordering. A fixed random seed is used to guarantee the reproducibility of the experimental results. The classifier, instantiated using XGBClassifier with the predefined hyperparameters, is consistently applied across all folds, ensuring coherence between the training and evaluation procedures.

3.3 Shapley Additive Explanations interpretability approach

SHAP method allows us to quantify the contribution of each variable to the model's predictions, both at the global level by assessing the importance of variables across the entire model, and at the local level, by explaining each prediction individually [32]. This dual interpretive capability proves particularly relevant in a pediatric monitoring context, where the precise identification of the factors triggering an alert is essential to avoid false alarms and limit the risk of misinterpretation. In this work, Interpretability is ensured by the TreeExplainer algorithm using tree structures to decompose predictions into additive Shapley contributions (φ_i). At the global level, the importance of variables is summarized using a summary bar plot, based on the sum of the absolute values of the contributions ($\Sigma|\varphi_i|$) across the test set, allowing predictors to be ranked according to their influence on the output probabilities. At the same time, local analysis via the Waterfall Plot measures, for each instance, the difference between the predicted probability and a reference value, where a positive contribution ($\varphi_i > 0$) indicates an effect favorable to the target class, while a negative contribution ($\varphi_i < 0$) indicates the opposite effect, thereby ensuring detailed and rigorous transparency of the model's decisions.

4. RESULTS AND DISCUSSION

4.1 Overall performance

The proposed multimodal model attained an accuracy of 0.93 and a macro F1-score of 0.90, signifying robust overall classification performance and balanced predictive capacity across classes. Receiver operating characteristic (ROC) analysis demonstrated significant differentiation between normal and abnormal classes, validating the efficacy of gradient boosting techniques for physiological anomaly identification. Figure 1 depicts the ROC curves generated during cross-validation. The curves continually approach the top-left corner of the ROC space, indicating a high true positive rate and a low false positive rate. The associated area under the curve (AUC) values demonstrate the strong discriminative capability of the model.

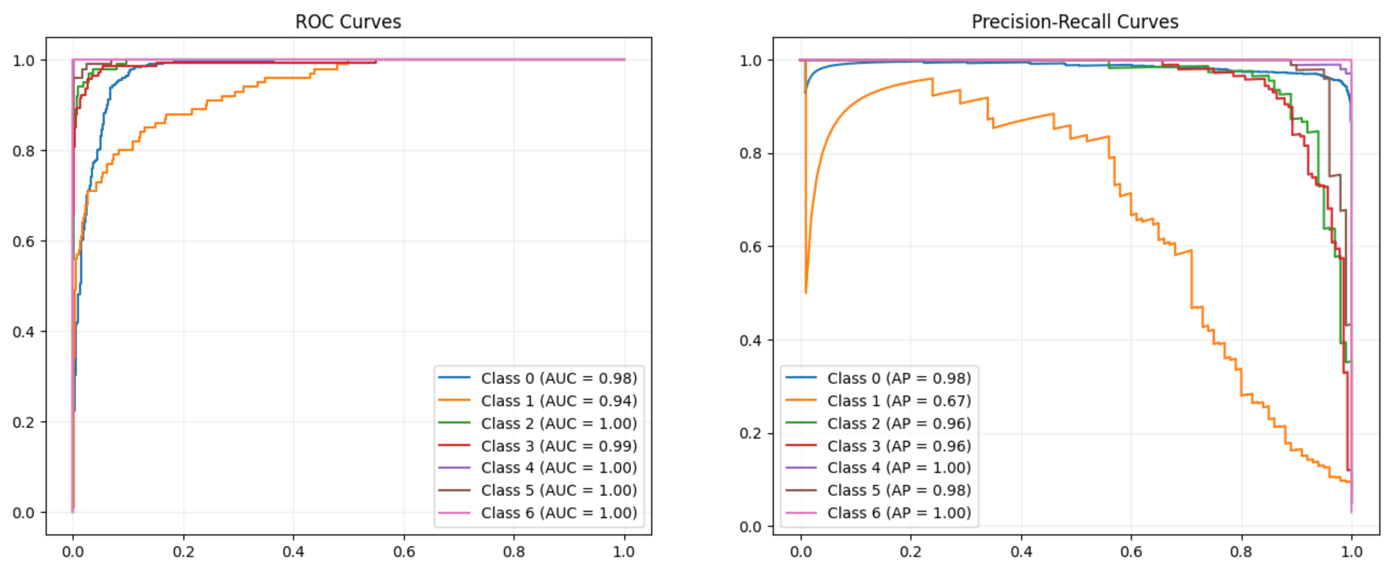

Figure 1. Receiver operating characteristic (ROC) precision-recall curves

4.2 Error analysis

4.2.1 Confusion matrix and error analysis

The confusion matrix, illustrated in Figure 2, shows high performance across several classes, indicating good discriminatory capability. In particular, the minority classes show remarkable results despite their low support, with perfect classification for class 6 (60/60) and a recall of 99% for class

4, indicating the learning of highly distinctive characteristic signatures. Similarly, classes 2, 3, and 5 exhibit high stability, with correct classification rates exceeding 90%, the remaining errors being marginal and predominantly skewed toward the dominant class (class 0). However, certain shortcomings persist, notably significant confusion between classes 0 and 1, with class 1 being the most affected: 27 cases of its samples are misclassified as class 0 (false negatives), while 51 samples from class 0 are predicted to belong to class 1 (false positives), resulting in a relatively low precision (0.58) for this category. A deeper inspection of the confusion matrix reveals a systematic error pattern between specific classes, which is further analyzed in the following section.

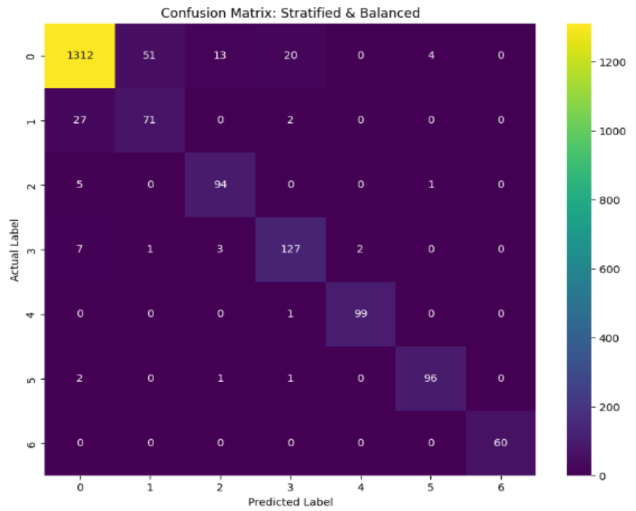

Figure 2. Confusion matrix

4.2.2 Improvement and comparative analysis

To further investigate the observed misclassification between classes 0 and 1, we conducted an in-depth error analysis of the feature space. The results indicated that the main source of confusion was related to inter-child physiological variability introduced in the synthetic dataset. The original feature set relied on six statistical descriptors of heart rate (minimum, maximum, mean, standard deviation, slope, and IQR. While these features captured temporal characteristics of the signal, they relied primarily on absolute physiological measurements and did not explicitly account for deviations relative to each child's baseline condition. Consequently, overlapping physiological ranges between classes reduced class separability and increased misclassification between classes with similar physiological characteristics.

To validate this hypothesis, we introduced two additional derived features, heart rate-delta and heart rate-ratio, designed to capture relative changes with respect to each child's baseline heart rate. These features explicitly model intra-subject variations and reduce the ambiguity introduced by inter-subject variability.

A comparative evaluation was then conducted between the baseline model and the enhanced feature representation. The results show a clear improvement in class separability, particularly for class 1, where misclassifications to class 0 decreased from 27 to 12 cases, representing a 55.6% reduction in false negatives and increasing recall to 88.0%. The resulting confusion matrix exhibits a more pronounced diagonal structure, with most classes achieving recall values above 95%. These findings indicate that the observed errors were primarily associated with the interaction between inter-child physiological variability and the limitations of the original feature representation rather than with the XGBoost classifier itself. The introduction of relative physiological indicators improved the modeling of inter-child variability and substantially enhanced class discrimination. This highlights the importance of incorporating relative physiological features when developing AI-based pediatric monitoring systems.

4.3 Significance analysis of predictors

4.3.1 Global explanation (model behavior)

Figure 3 provides a detailed understanding of the classifier's decision-making mechanisms. First, it highlights the dominant role of the IMU_motion_level variable, which appears to be the most influential predictor, particularly for class 6 (orange), indicating that the level of activity or movement is a major discriminating factor for this category. Second, physiological parameters, particularly oxygen saturation measurements (SpO2_mean, SpO2_min) and heart rate (HR) measurements (HR_mean, HR_min), play a significant role, with SpO_2 proving decisive for classes 4 (light blue) and 3 (magenta). Furthermore, variables associated with temperature and the PPG signal contribute in a more targeted manner to the prediction of classes 5 (pink) and 1 (green). However, the relatively low contribution of PPG characteristics for class 1 may help explain the more limited classification performance observed for this category.

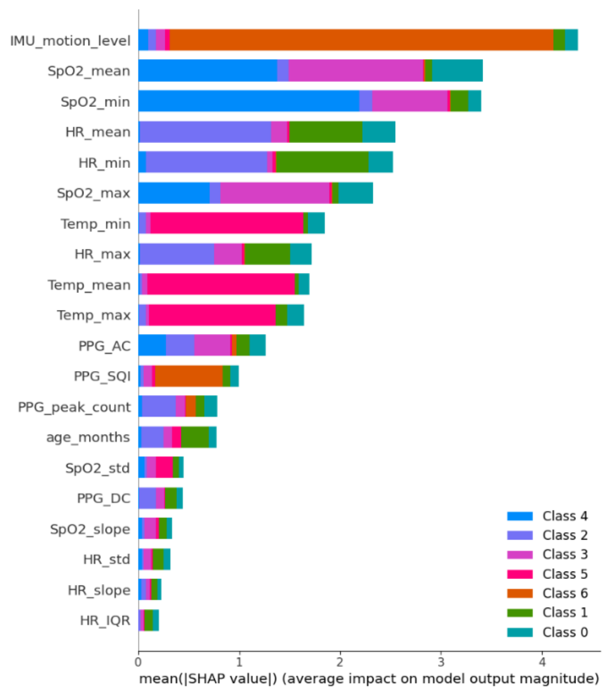

Figure 3. Shapley Additive Explanations (SHAP) summary bar plot

4.3.2 Local explanation (individual case)

SHAP-based local explainability is a key element in overcoming the "black box" nature of the model by providing a transparent justification for each prediction at the individual level, particularly by breaking down the precise influence of each vital sign on the final score, which ensures the clinical traceability essential for validating algorithmic decisions prior to any medical intervention. To this end, prediction tests were conducted using simulated data for several patients (Appendix A). Analysis of the SHAP values highlights distinct biometric contribution profiles across clinical classes, thereby revealing the heterogeneity of the physiological determinants involved in predicting pediatric health states.

The findings indicate that, for the case PED-99 (Figure 4), the categorization into Class 6 (movement) is predominantly elucidated by a significant contribution from the level of physical activity, wherein the variable IMU_motion_level = 2 serves as a decisive factor with an effect of +7.51 on the overall score, signifying that motor agitation is the principal distinguishing biomarker. This analysis is corroborated by an elevated SQI (PPG_SQI = 0.95), discounting the likelihood of artifacts, alongside the relative consistency of physiological indices such as SpO2_min (94%) and HR_std (8), implying the absence of concomitant clinical distress. In the case of the PED-91 profile (Class 1), the forecast is chiefly influenced by cardiac dynamics, with the minimum, maximum, and average heart rate metrics collectively and positively affecting the score fit (Figure 5). Lastly, the PED-88 instance (Class 5) is marked by a predominance of thermal variables, with the minimum and maximum temperatures serving as the principal explanatory elements (+3.5 and +2.51), relegating cardiac and oxygenation variables to a secondary role (Figure 6).

This variability in the dominant contributions—ranging from motion to cardiometry and then to thermometry—highlights the model's ability to capture physiological signatures specific to each class, while confirming the robustness of a multidimensional approach that integrates inertial and biometric sensors.

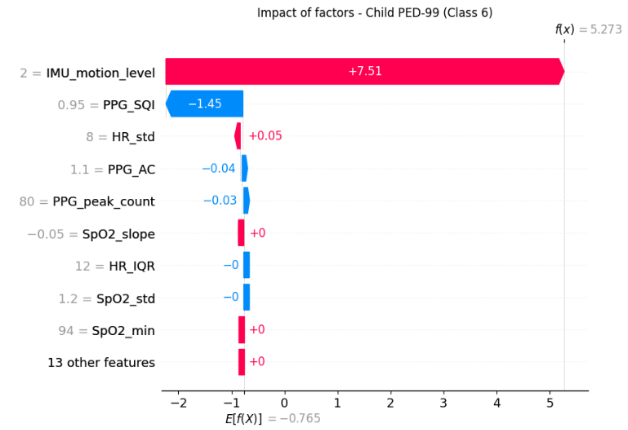

Figure 4. Impact of factors - Child PED-99

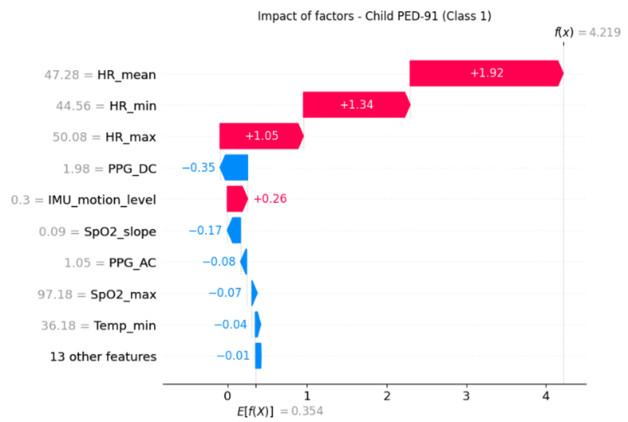

Figure 5. Impact of factors - Child PED-91

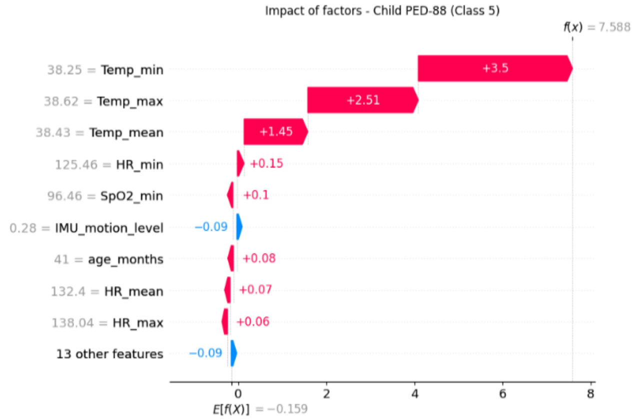

Figure 6. Impacts of factors - Child PED-88

4.3.3 Clinical validation of interpretability

The examination of the global SHAP plot reveals a significant correlation between the variable importance identified by the model and the tenets of pediatric clinical semiology. This implies that the model is based on both statistical correlations and established physiological principles. It is important to note that oxygenation-related variables like SpO2_mean and SpO2_min are important for finding classes related to respiratory distress (Classes 3 and 4). This is because oxygen saturation is a key biomarker for hypoxia in clinical practice. Moreover, cardiac dynamics, indicated by HR_mean and HR_min, are crucial in categorizing cases of hemodynamic instability (Classes 1 and 2), illustrating the dependence of pediatric cardiac output on heart rate—a critical element in diagnosing rhythm disorders such as bradycardia and tachycardia. The significant influence of thermal variables (Temp_min, Temp_max) for Class 5 highlights the model's integration of thermal homeostasis mechanisms, essential for detecting infectious states or the risks of neonatal hypothermia. Finally, the importance of motion level (IMU_motion_level) and signal quality (PPG_SQI), particularly for Class 6, demonstrates that contextual factors are integrated into the model's decision process. The model combines motor activity, which is an important sign in pediatric assessments, with the reliability of biometric measurements. This ensures consistent interpretation of vital signs under real-world monitoring conditions.

4.3.4 Consistency of key features

In order to evaluate the robustness of the model, the stability of the features was analyzed through cross-validation, utilizing the Jaccard index and Spearman's correlation coefficient. The results show excellent structural consistency of the model, with an average Jaccard index of 0.76 and a Spearman correlation of 0.96, indicating that the order of importance of the variables remains virtually identical from one sample to another. This robustness is confirmed by the identification of a stable core of four key variables (HR_mean, IMU_motion_level, SpO2_mean, and SpO2_min) exhibiting a perfect stability ratio of 1.0 (Figure 7). The convergence of these metrics illustrates that the model's predictions are not influenced by random variations in the training data, but are based on robust and reproducible physiological descriptors, thereby ensuring the system's reliability in a multi-class context.

Because SHAP explanations are derived from the trained

XGBoost model, the high consistency of feature rankings across cross-validation folds provides evidence that the generated SHAP explanations are stable and reproducible rather than being artifacts of a particular training split.

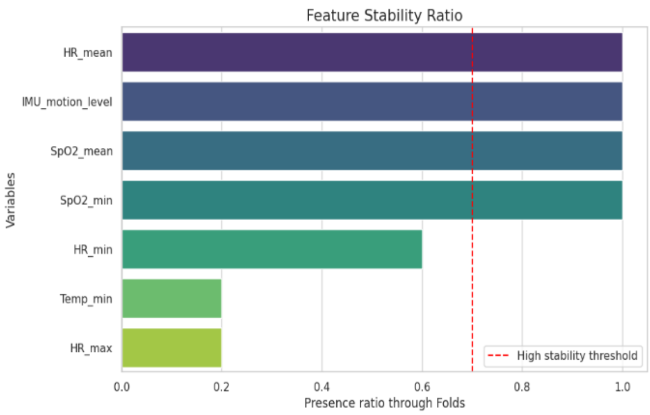

Figure 7. Feature stability ratio

4.4 Robustness analysis

4.4.1 Stratified cross-validation and model robustness

To assess the model's stability across different subsets of data, the model underwent five-fold stratified cross-validation, which ensured that the class distribution was preserved and provided a robust estimate of generalization performance. The results indicate an average accuracy of 95.15%, associated with a low standard deviation of 0.13% and a 95% confidence interval ranging from 94.99% to 95.31%, reflecting limited dispersion around the mean. The relative variance, also 0.13%, remains well below the commonly accepted 3% threshold for good stability, thus confirming the high consistency of performance across the different training and test partitions, as well as low sensitivity to data partitioning. Table 4 shows that the model exhibits high and stable performance during the training phase, with an accuracy of 0.988 ± 0.001 and an F1-score of 0.987 ± 0.001, reflecting high discriminatory power and minimal variability between folds. In cross-validation, a moderate decrease is observed (accuracy = 0.950 ± 0.004; F1-score = 0.946 ± 0.004), suggesting good generalization despite slight overfitting. On the independent test set, performance reached an accuracy of 0.929 and an F1-score of 0.932, indicating a gradual but controlled decline between training and final evaluation. The overall discrepancy of approximately 6% between the training and testing phases remains consistent with a robust model that shows no significant overfitting, thereby confirming its ability to maintain reliable predictive performance under independent conditions—an essential requirement for practical applications in the detection of physiological abnormalities.

Table 4. Summary of model performance across cross-validation and independent testing

Metric	Training (Cross-Validation)	Validation (Cross-Validation)	Test (Independent)
Accuracy	0.988 ± 0.001	0.950 ± 0.004	0.929
F1-score	0.987 ± 0.001	0.946 ± 0.004	0.932

4.4.2 Sensitivity analysis

To analyze the impact of changes in hyperparameters (learning rate, number of trees, maximum depth) on performance, a realistic range of values was defined for each parameter. The sensitivity analysis shows that the model remains stable for moderate variations in hyperparameters: learning_rate 0.03–0.2, maximum depth 5–6, and number of trees ≥300. Performance is slightly reduced when the learning rate is too low, or the trees are too shallow, whereas beyond these values, accuracy and the F1-score stabilize (accuracy ≈ 92.9–93.5%, F1-score ≈ 0.931–0.936) (Figure 8). These results confirm that the currently used parameters (learning rate = 0.03, maximum depth = 5, number of trees = 500) offer an optimal trade-off between accuracy, stability, and robustness to parameter variations.

4.4.3 Prediction stability

To evaluate the model's local stability in the face of minor physiological variations and sensor noise, proportional Gaussian noise was injected into the normalized features of the test dataset, with 30 repetitions per perturbation level. The prediction consistency rate, defined as the proportion of samples retaining the same class, was measured for perturbations of ±1%, ±2%, ±3%, and ±5%, reaching 94.04%, 83.08%, 72.37%, and 98.57%, with very low standard deviations (0.44–0.17%), before and after data normalization. These results indicate that the model remains robust for small variations (±1%) and retains strong stability at ±5% after normalization, while classes near the decision boundaries exhibit increased sensitivity to moderate perturbations (±2–3%) in the absence of normalization. Thus, this analysis highlights the critical role of feature normalization in improving the model's local robustness, ensuring that realistic physiological fluctuations and sensor noise do not significantly alter the predictions.

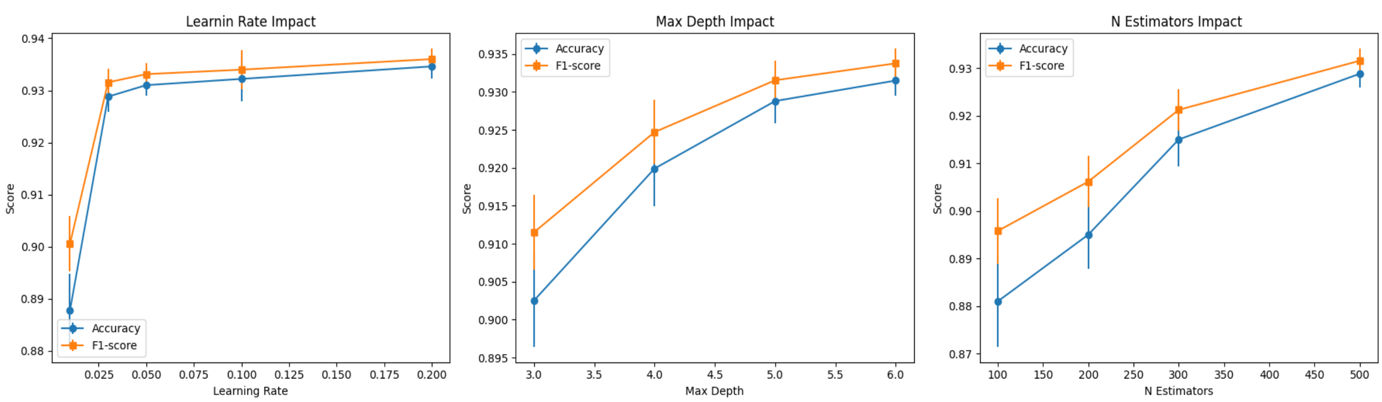

Figure 8. Learning rate-max depth-*n* estimators impact

4.5 Comparison with other machine learning models

4.5.1 Detailed methodological benchmarking

To provide a more comprehensive evaluation of the proposed approach, an extended comparative study was conducted including both classical machine learning models and recent state-of-the-art gradient boosting methods. The evaluated models include Logistic Regression, SVM, Random Forest, XGBoost, as well as modern ensemble approaches such as gradient boosting, LightGBM, and CatBoost.

The results presented in Table 5 indicate that all advanced tree-based ensemble methods significantly outperform linear models and SVM in terms of both Accuracy and macro F1-score. Logistic Regression and SVM achieve the lowest performance (Accuracy: 0.84 and 0.87, respectively), highlighting their limited capacity to capture nonlinear relationships in the physiological feature space. Random Forest provides a stronger baseline (Accuracy: 0.89, F1-macro: 0.87), but remains inferior to boosting-based approaches.

Among the boosting family, gradient boosting achieves the highest overall performance (Accuracy: 0.9511, F1-macro: 0.9126), closely followed by CatBoost (Accuracy: 0.9501, F1-macro: 0.9101) and LightGBM (Accuracy: 0.9441, F1-macro: 0.9101). XGBoost also demonstrates competitive performance (Accuracy: 0.9314, F1-macro: 0.9008), although slightly lower than the best-performing gradient boosting variants in this experimental setting.

Beyond predictive performance, computational efficiency presents an important distinction between models. While traditional gradient boosting exhibits significantly higher training time (approximately 365 seconds), LightGBM, CatBoost, and XGBoost achieve comparable performance with substantially lower computational cost (approximately 5–7 seconds), making them more suitable for resource-constrained or real-time monitoring scenarios.

Overall, although gradient boosting achieves marginally superior accuracy in this dataset configuration, the differences between the top-performing boosting models remain relatively small. Therefore, model selection should not rely solely on predictive performance but also consider computational efficiency and interpretability. In this context, XGBoost was retained in the proposed framework due to its well-established balance between performance, training efficiency, and compatibility with SHAP-based explainability, which is a central requirement for the decision-support system.

Finally, the consistency of results across multiple boosting methods further confirms that nonlinear ensemble learning is particularly well-suited for modeling complex physiological interactions in pediatric monitoring applications.

Table 5. Model accuracy comparison

Model	Accuracy	F1-Score	Training Time (s)
XGBoost	0.93	0.90	5.60
Random Forest	0.88	0.87	9.14
Support vector machine (SVM)	0.87	0.85	6.54
Logistic Regression	0.83	0.84	0.49
Gradient boosting	0.95	0.91	365.07
LightGBM	0.94	0.91	5.02
CatBoost	0.95	0.91	6.84

4.6 Ablation analysis

To assess the impact of the modality, the model was trained on various combinations of sensors to quantify the informational contribution of each source. The results highlight the superiority of the multimodal approach, which achieves an average cross-validation accuracy of 94.92% and a test set accuracy of 0.93, outperforming PPG alone by approximately 5% and temperature and IMU modalities considered individually by more than 20% (Table 6). This significant improvement demonstrates that the fusion of physiological, thermal, and contextual signals enables the capture of more comprehensive and discriminative patterns than any single modality, making multimodality a complementary integration of heterogeneous information that enhances the robustness and generalization capability of the detection system.

Table 6. Multimodal vs. monomodal performance

Modality	Accuracy
Multimodal	0.93
Photoplethysmography (PPG) only	0.88
Temperature only	0.74
Motion only	0.72

4.7 Discussion of limitations

Despite the promising results obtained in this study, several limitations should be acknowledged when interpreting the findings.

From a theoretical perspective, the results of this study contribute to the growing body of evidence supporting the use of gradient boosting methods combined with SHAP-based explainability for multimodal physiological signal analysis. More specifically, the introduction of relative physiological indicators, such as heart rate-delta and heart rate-ratio, demonstrates that intra-subject deviation features can substantially improve class separability in heterogeneous pediatric populations, a finding with broader implications for the design of personalized monitoring frameworks. Furthermore, the quantitative validation of SHAP consistency across cross-validation folds establishes a methodological precedent for interpretability assessment in non-clinical AI-based monitoring systems, where ground-truth clinical annotations are not available.

First, the evaluation was conducted exclusively on a physiologically informed synthetic dataset. Although the data generation process was designed to incorporate age-dependent physiological variability, clinically plausible relationships among vital signs, and realistic temporal fluctuations, synthetic data cannot fully reproduce the complexity and variability of real pediatric physiological measurements. Consequently, the reported performance should be interpreted as evidence of the effectiveness of the proposed framework within a controlled simulation environment rather than as proof of clinical validity. Validation using real-world pediatric datasets and prospective monitoring scenarios remains necessary to assess the generalizability and practical applicability of the proposed approach.

Second, the present work focuses primarily on the development and evaluation of the intelligent analytics and decision-support layer of the framework. While the proposed architecture is intended for future integration within an IoT-based monitoring system, no real-world deployment was performed in this study. Therefore, important operational factors associated with continuous monitoring in childcare environments, such as communication reliability, latency, energy consumption, device interoperability, and long-term system stability, were not experimentally assessed.

Third, the robustness evaluation was limited to Gaussian noise perturbations. Although this analysis provides an initial assessment of model resilience to measurement uncertainty, it does not capture several challenges commonly encountered in real-world IoT monitoring systems. Issues such as sensor displacement, temporary sensor detachment, signal degradation, packet loss, synchronization errors, communication interruptions, and prolonged periods of missing data were not explicitly modeled. Evaluating the impact of these factors will be an important step toward validating the reliability of the complete system under realistic operating conditions.

Finally, XGBoost was selected in this study because it offers a favorable balance between predictive performance, computational efficiency, and explainability when combined with SHAP-based interpretation. However, this choice should not be interpreted as evidence that XGBoost is necessarily the optimal model for embedded or edge deployment scenarios. The objective of this work was to establish and validate an explainable anomaly detection framework rather than to perform an exhaustive comparison of deployment-oriented machine learning models. Future studies should investigate alternative lightweight architectures and conduct detailed benchmarking on resource-constrained platforms to identify the most suitable solutions for real-time pediatric monitoring applications.

To progressively address these limitations, and as the next stage of this research, a real-world validation study is planned. The target population will consist of children aged 1 to 5 years monitored in daycare or nursery settings, with a planned sample size of approximately 100 to 150 participants, consistent with the scope of feasibility studies reported in the pediatric wearable monitoring literature. Data will be collected using non-invasive wearable sensors measuring heart rate, SpO_2, body temperature, PPG signals, and motion data, in direct alignment with the modalities evaluated in the present study. Prior to any data collection, ethical approval will be obtained from the relevant institutional review board, and informed written consent will be required from the legal guardians of all participating children. All data will be anonymized and processed in accordance with applicable data protection regulations. The evaluation protocol will involve comparison of model predictions against annotations provided by qualified pediatric nurses or physicians, enabling both quantitative performance assessment and clinical validation of the SHAP-based explanations. Model performance will be reported using the same metrics employed in the present study—accuracy and macro F1-score—to ensure direct comparability between the synthetic and real-world evaluation phases.

These limitations define important directions for future research and represent the next stages in the progressive development and real-world validation of the framework.

5. CONCLUSION

This work proposed a multimodal physiological anomaly detection framework for pediatric monitoring in daycare settings, integrating heart rate, SpO_2, body temperature, PPG morphology, and IMU data within an interpretable XGBoost-based classifier enhanced by SHAP explainability. The results obtained on a physiologically informed synthetic dataset demonstrate robust detection performance, with a test accuracy of 0.93 and a macro F1-score of 0.90 across seven physiological classes, confirming the effectiveness of the proposed approach within a controlled simulation environment.

The principal contributions of this work include the multimodal fusion of heterogeneous physiological signals, the introduction of child-specific relative features (heart rate-delta, heart rate-ratio) to reduce inter-individual variability, and the quantitative validation of SHAP consistency across cross-validation folds. The lightweight nature of XGBoost further supports its suitability for future deployment in resource-constrained IoT environments.

Nevertheless, the reliance on synthetic data limits the generalizability of the findings, and real-world clinical validation remains a necessary next step, as outlined in the planned validation roadmap. The proposed framework represents a promising and reproducible foundation for future IoT-enabled pediatric monitoring systems, pending validation on real-world data and deployment studies.

REFERENCES

[1] Deshmukh, S., Shah, S., Wahedna, A., Sabnis, N. (2025). IoT-enabled smart healthcare system with machine learning for real-time vital sign monitoring and anomaly detection. Indonesian Journal of Electrical Engineering and Computer Science, 39(2): 1155-1163. https://doi.org/10.11591/ijeecs.v39.i2.pp1155-1163

[2] Sunny, J.S., Patro, C.P.K., Karnani, K., et al. (2022). Anomaly detection framework for wearables data: A perspective review on data concepts, data analysis algorithms and prospects. Sensors, 22(3): 756. https://doi.org/10.3390/s22030756

[3] Mellal, N., Maatallah, S.H., Merazga, A., Bouchouareb, R., Nacer, S. (2026). IoT-enabled connected incubator with redundant communication for real-time neonatal monitoring. Indonesian Journal of Electrical Engineering and Computer Science, 41(2): 633-644. https://doi.org/10.11591/ijeecs.v41.i2.pp633-644

[4] Pavithra, D., Parameswaran, T., Choudhry, M.D., Amrutha, S., Deepa, T., Kiruthiga, R. (2024). Application of LSTM networks for continuous patient monitoring and anomaly detection in wearable health devices. Indian Journal of Science and Technology, 17(37): 3909-3921. https://doi.org/10.17485/ijst/v17i37.2600

[5] Loh, H.W., Ooi, C.P., Seoni, S., Barua, P.D., Molinari, F., Acharya, U.R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Computer Methods and Programs in Biomedicine, 226: 107161. https://doi.org/10.1016/j.cmpb.2022.107161

[6] Yang, A.C., Ma, W.M., Chiang, D.H., et al. (2025). Early prediction of sepsis using an XGBoost model with single time-point non-invasive vital signs and its correlation with C-reactive protein and procalcitonin: A multi-center study. Intelligence-Based Medicine, 11: 100242. https://doi.org/10.1016/j.ibmed.2025.100242

[7] Chen, T.Q., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, pp. 785-794. https://doi.org/10.1145/2939672.2939785

[8] Divyabharathi, S., Madhavan, S. (2025). An explainable IoT-based framework for anomaly detection and emergency decision management in smart healthcare. In 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, pp. 678-682. https://doi.org/10.1109/icscds65426.2025.11167637

[9] Lin, S.H., Fu, B., Wang, W., Liu, M., Sun, X. (2022). Dynamic sepsis prediction for intensive care unit patients using XGBoost-based model with novel time-dependent features. IEEE Journal of Biomedical and Health Informatics, 26(8): 4258-4269. https://doi.org/10.1109/JBHI.2022.3171673

[10] Mayer, L., Rasmussen, S.M., Molgaard, J., Gu, Y., Aasvang, E.K., Meyhoff, C.S. (2022). Prediction of serious adverse events from nighttime vital signs values. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, pp. 2631-2634. https://doi.org/10.1109/EMBC48229.2022.9871778

[11] Shah, S. (2012). Vital sign monitoring and data fusion for paediatric triage. Doctoral thesis, Oxford University, UK. https://ora.ox.ac.uk/objects/uuid:80ae66e3-849b-4df1-b064-f9eb7530200d.

[12] Fleming, S. (2010). Measurement and fusion of non-invasive vital signs for routine triage of acute paediatric illness. Doctoral thesis, Oxford University, UK. https://ora.ox.ac.uk/objects/uuid:840d94b0-041f-4b15-8b37-c2e37c999f3e.

[13] Gangadharan, V. (2013). Automated multi-parameter monitoring of neo-nates. Doctoral thesis, University College London. https://discovery.ucl.ac.uk/id/eprint/1417079/.

[14] Band, S.S., Yarahmadi, A., Hsu, C.C., et al. (2023). Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. Informatics in Medicine Unlocked, 40: 101286. https://doi.org/10.1016/j.imu.2023.101286

[15] Kumar, R., Srirama, V., Chadaga, K., Muralikrishna, H., Sampathila, N., Prabhu, S. (2024). Using explainable machine learning methods to predict the survivability rate of pediatric respiratory diseases. IEEE Access, 12: 189515-189534. https://doi.org/10.1109/access.2024.3516045

[16] Kapsner, L.A., Feißt, M., Purbojo, A., et al. (2024). Using machine learning and feature importance to identify risk factors for mortality in pediatric heart surgery. Diagnostics, 14(22): 2587. https://doi.org/10.3390/diagnostics14222587

[17] Zhou, J.T., Xie, Y.J., Liu, Y., et al. (2025). Interpretable machine learning model for early prediction of disseminated intravascular coagulation in critically ill children. Scientific Reports, 15: 11217. https://doi.org/10.1038/s41598-025-91434-w

[18] Prathyanga, A., Shyaminda, P., Chamikara, P., Lakshan, S., Thelijjagoda, S., Kasthurirathna, D. (2024). Intelligent daycare: Enhancing child safety with IoT and machine learning innovations. In 2024 9th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 530-538. https://doi.org/10.1109/ICCES63552.2024.10859472

[19] Basak, A., Narasimhan, S., Bhunia, S. (2011). KiMS: Kids' health monitoring system at day-care centers using wearable sensors and vocabulary-based acoustic signal processing. In 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services, Columbia, MO, USA, pp. 1-8. https://doi.org/10.1109/HEALTH.2011.6026744

[20] Abo-Zahhad, M., Sayed, M.S., Abd El-Malek, A.H., Hossam, M., Atef, N., Reda, S. (2022). Intelligent IoT-based tracking system for constant supervision of health status and daily activities of children. In 2022 10th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC), Alexandria, Egypt, pp. 175-180. https://doi.org/10.1109/JAC-ECC56395.2022.10044030

[21] Pradeepa, M., Jamberi, K., Sajith, S., Rama Bai, M., Prakash, A., Kazi, K.S.L. (2022). Student health detection using a machine learning approach and IoT. In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Mysuru, India, pp. 1-5. https://doi.org/10.1109/MysuruCon55714.2022.9972445

[22] Fleming, S., Thompson, M., Stevens, R., et al. (2011). Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. The Lancet, 377(9770): 1011-1018. https://doi.org/10.1016/S0140-6736(10)62226-X

[23] World Health Organization. (2013). Pocket Book of Hospital Care for Children: Guidelines for the Management of Common Childhood Illnesses. Geneva, Switzerland. https://www.who.int/publications/i/item/9789241548373.

[24] Topjian, A.A., Raymond, T.T., Atkins, D., et al. (2020). Part 4: Pediatric basic and advanced life support: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation, 142(16_suppl_2): S469-S523. https://doi.org/10.1161/CIR.0000000000000901

[25] American Academy of Pediatrics. (2021). Fever and Your Child. Pediatric Patient Education. https://doi.org/10.1542/peo_document040

[26] Ergün, S. (2023). Explaining XGBoost predictions with SHAP value: A comprehensive guide to interpreting decision tree-based models. New Trends in Computer Sciences, 1(1): 19-31. https://doi.org/10.3846/ntcs.2023.17901

[27] Alizamir, M., Wang, M., Ikram, R.M.A., et al. (2025). An interpretable XGBoost-SHAP machine learning model for reliable prediction of mechanical properties in waste foundry sand-based eco-friendly concrete. Results in Engineering, 25: 104307. https://doi.org/10.1016/j.rineng.2025.104307

[28] Sekadakis, M., Garefalakis, T., Moertl, P., Yannis, G. (2026). Analyzing SHAP values of XGBoost algorithms to understand driving features affecting take-over time from vehicle alert to driver action. Displays, 91: 103263. https://doi.org/10.1016/j.displa.2025.103263

[29] Shwartz-Ziv, R., Armon, A. (2022). Tabular data: Deep learning is not all you need. Information Fusion, 81: 84-90. https://doi.org/10.1016/j.inffus.2021.11.011

[30] Zhou, L.L., Zhu, Q., Chen, Q., Wang, P., Huang, H. (2025). Predicting hospital outpatient volume using XGBoost: A machine learning approach. Scientific Reports, 15: 17028. https://doi.org/10.1038/s41598-025-01265-y

[31] Huang, K.Y., Hsu, Y.L., Chung, C.L., et al. (2025). Enhancing healthcare AI stability with edge computing and machine learning for extubation prediction. Scientific Reports, 15: 17858. https://doi.org/10.1038/s41598-025-02317-z

[32] Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, pp. 4768-4777. https://dl.acm.org/doi/10.5555/3295222.3295230

APPENDIX

A. Representative test data for three scenarios

Input_data = {
'child_id' : "PED-99",'age_months' : 24, 'HR_min' : 70, 'HR_max' : 110, 'HR_mean' : 85, 'HR_std' : 8, 'HR_slope' : 0.2, 'HR_IQR' : 12,'SpO2_min' : 94, 'SpO2_max' : 99, 'SpO2_mean' : 97, 'SpO2_std' : 1.2, 'SpO2_slope' : -0.05, 'SpO2_dip_count' : 1,'Temp_min' : 36.2, 'Temp_max' : 37.5, 'Temp_mean' : 36.8, 'PPG_AC' : 1.1, 'PPG_DC' : 0.6, 'PPG_peak_count' : 80,'PPG_SQI' : 0.95, 'IMU_motion_level' : 2, 'Signal_quality_flag' : 1
}
Input_data1 = {
'child_id' : "PED-91", 'age_months' : 58, 'HR_min' : 44.56, 'HR_max' : 50.08, 'HR_mean' : 47.28, 'HR_std' : 1.96, 'HR_slope' : 0.37, 'HR_IQR' : 3.03, 'SpO2_min' : 95.78, 'SpO2_max' : 97.18, 'SpO2_mean' : 96.39, 'SpO2_std' : 0.40, 'SpO2_slope' : 0.09, 'SpO2_dip_count' : 0, 'Temp_min' : 36.18, 'Temp_max' : 36.57,'Temp_mean' : 36.33, 'PPG_AC' : 1.05, 'PPG_DC' : 1.98, 'PPG_peak_count' : 9, 'PPG_SQI' : 0.87, 'IMU_motion_level' : 0.30, 'Signal_quality_flag' : 1
}
Input_data2 = {
'child_id' : "PED-88", 'age_months' : 41, 'HR_min' : 125.46, 'HR_max' : 138.04, 'HR_mean' : 132.40, 'HR_std' : 3.91, 'HR_slope' : 0.003, 'HR_IQR' : 4.65, 'SpO2_min' : 96.46, 'SpO2_max' : 97.61, 'SpO2_mean' : 97.02, 'SpO2_std' : 0.38, 'SpO2_slope' : 0.004, 'SpO2_dip_count' : 0, 'Temp_min' : 38.25, 'Temp_max' : 38.62, 'Temp_mean' : 38.43, 'PPG_AC' : 0.86, 'PPG_DC' : 1.23, 'PPG_peak_count' : 21, 'PPG_SQI' : 0.86, 'IMU_motion_level' : 0.28, 'Signal_quality_flag' : 1
}