Received: 22 November 2025 | Revised: 3 April 2026 | Accepted: 3 June 2026 | Published online: 29 June 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–22
DOI: 10.47852/bonviewAIA62028365

BON VIEW PUBLISHING

Explainable AI-Driven Clinical Decision Support for Mortality Risk Stratification in Pediatric Respiratory Disorders

Khandaker Mohammad Mohi Uddin[1,*], Anjuman Ansary[1], Nazim Uddin[2] and Md. Tofael Ahmed Bhuiyan[1]

[1] *Department of Computer Science and Engineering, Southeast University, Bangladesh*

[2] *Department of Computer Science and Engineering, Chandpur Science and Technology University, Bangladesh*

Abstract: Nowadays, respiratory illnesses are common among children around the world, which is contributing significantly to child mortality due to rapid growth. These diseases appear with several symptoms, and several may share similar symptoms; the most well-known include asthma, bronchiolitis, and pneumonia. The mortality rate associated with chronic pediatric respiratory conditions is increasing annually, underscoring the need to assess the severity of these diseases. This research aims to improve the mortality prediction in pediatric respiratory disorders by using a robust machine learning architecture and a comprehensive dataset from the Children's Hospital of Rabat, Morocco. Ten machine learning algorithms were implemented after thorough preprocessing, which included encoding, feature selection, oversampling, and hyperparameter tuning. Random Forest, Logistic Regression, AdaBoost, Multilayer Perceptron, K-Nearest Neighbors, Decision Trees, CatBoost, XGBoost, Naïve Bayes, and Gradient Boosting have been applied to get better results. With an accuracy rate of 99.29%, a precision of 92.68%, a recall rate of 92.55%, and an F1-score of 92.58%, Random Forest performed more effectively than the other methods. Explainable artificial intelligence techniques, such as Shapley Additive Values and Local Interpretable Model-agnostic Explanations, ensure clinician trust and model interpretability.

Keywords: pediatric respiratory, SMOTE-ENN, ensemble machine learning, hyperparameter tuning, explainable AI

1. Introduction

Currently, chronic respiratory conditions such as asthma, bronchospasm, influenza, laryngeal dyspnea, bronchiolitis, and staphylococcal pneumonia are a major public health concern [1]. It is now regarded as a significant health burden and the fourth leading cause of death. These diseases are more common among children. These respiratory disorders have become more prevalent due to increasing air pollution. As children breathe orally, many harmful substances enter their lower airways, which can lead to respiratory disorders [2, 3].

Therefore, the survivability rate for respiratory infections in pediatric intensive care units (PICUs) has a substantial effect on children's lives as well as the healthcare system. Nearly 40% of PICU hospitalizations are caused by pediatric respiratory conditions like acute pneumonia, Acute Respiratory Distress Syndrome (ARDS), and respiratory failure, which have a 7–15% fatality risk [4]. Alongside a concerning drop in the neonatal survival rate, pediatric mortality is slowly getting worse every day [5]. In the PICU, survivors of severe respiratory conditions frequently have long-term repercussions, including psychological problems, physical disabilities, and neurodevelopmental difficulties. Six months following discharge, new functional difficulties were reported by about 25% of juvenile ARDS survivors [6]. Children with severe respiratory illness are frequently admitted to the PICU, resulting in significant healthcare costs associated with extended hospital stays, intensive care, and high-technology respiratory support [7–10]. According to a study by the Pan American Health Organization and World Health Organization, the inexperience of junior doctors in primary clinics, as well as children under two who are unable to communicate their pain clearly and understandably based on parental complaints, is a major problem in pediatrics that makes it difficult to diagnose respiratory diseases accurately. In this case, they must use chest radiography to obtain an accurate medical diagnosis, which is costly and unavailable to two-thirds of the world's population [11]. Finding available and affordable alternative options will therefore be the focus of the effort and orientation. Therefore, to take the exact treatment promptly, it's important to know the severity of infections, which can be determined from the survivability rate. In order to protect young lives, enable focused interventions, and allocate resources to reduce catastrophic consequences, it is crucial to predict pediatric mortality [12]. Significant medical resources, such as ventilators, specialist drugs, and qualified healthcare professionals, are needed to manage critically ill children with respiratory diseases. This can put a burden on the healthcare system, possibly resulting in shortages and higher expenses [13, 14]. Families, caregivers, and

***Corresponding author:** Khandaker Mohammad Mohi Uddin, Department of Computer Science and Engineering, Southeast University, Bangladesh. E-mail: mohiuddin.kh@seu.edu.bd

medical professionals are all affected emotionally and psychologically when a child dies in the PICU from a respiratory illness, which can result in long-term sorrow and mental health issues. To lessen these effects, early discovery, effective management, and technological breakthroughs are crucial.

In order to find speedy, accurate answers at the lowest possible cost, the world is focused on using and introducing current technology in medicine. Since machine learning (ML) techniques are thought to be a cutting-edge and contemporary means of relieving the strain on healthcare systems through quick and precise diagnosis and lowering the need for pricey equipment and devices, they have attracted a lot of attention in recent years for the detection of respiratory diseases using precise algorithms and models [15, 16]. We may endeavor to lessen the impact of these illnesses on children's lives and relieve the strain on the healthcare system by funding research, medical resources, and preventative measures [17, 18].

The fast development of ML and artificial intelligence (AI) makes the combination with medical science inevitable. ML algorithms have proven to be excellent at predicting a number of ailments. As a subset of AI, machine learning (ML) enables stakeholders to select the optimal characteristics to gain the least amount of uncertainty and extract important information from the data that is available. Models have developed several clinical decision-support systems and forecasted health measures by utilizing ML algorithms [19]. One of the subfields of data science and related fields is ML, which enables computer models to learn mathematical rules by repeating their data in context without needing to be explicitly programmed to develop or enhance the data that has been gathered [20], use clinical data to predict respiratory conditions, for instance. Conversely, traditional modeling refers to data that is directly designed and based on a predefined field. Big data availability and computer processing speed have increased the use of ML applications, particularly in clinical settings where ML techniques based on static or well-established statistical data with multiple variables are being adopted [21]. Several features of ML applications were found to aid in the early detection of childhood asthma. The variety of data sources is what improves the prediction's accuracy since it can create and raise the likelihood of precise predictive scores for pediatric respiratory disorders in comparison to conventional models. ML can be classified as supervised, unsupervised, or reinforcement learning based on the model's structure.

In order to make AI more transparent and employ it in delicate areas where the patient and doctor can comprehend the results without knowing anything about the model, explainable AI (XAI) has been proposed [22]. XAI methods such as Explain Like I'm 5 (ELI5), QLattice, Local Interpretable Model-agnostic Explanations (LIME), and Shapley Additive Values (SHAP) can be used to comprehend the ML model predictions. Gaining the trust of people or organizations can be accomplished by outlining the characteristics and explanations for the AI's output [23]. AI-based findings have the potential to improve diagnosis accuracy, benefiting patients and saving physicians' time [24]. By combining the theoretical understanding of physicians with the analysis of large amounts of real-world data, diagnoses can become more accurate and efficient. However, the problem might not be fully resolved by depending only on ML and classification methods. Establishing a strong framework for acquiring information and making sure that all requirements are fulfilled is also essential.

Specifically, survival rates in pediatric critical care units are predicted using ML, which is advanced in this study. The important points include:

1) Extensive data preparation involves converting categorical data, handling incomplete data, adjusting the scale of features, eliminating extreme values, selecting relevant features, and using oversampling to correct imbalances in the data categories.
2) We assessed 10 ML techniques, including Random Forest, Logistic Regression, AdaBoost, Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), Decision Trees, CatBoost, XGBoost, Naïve Bayes, and Gradient Boosting.
3) The study showed that Random Forest produced the best results. Random Forest showed excellent performance, with 99.29% accuracy, 92.68% precision, 92.55% recall, and a 92.58% F1-score.
4) By using the SHAP model and explainable AI, we improved interpretability, which increased confidence in the AI's conclusions.
5) Fine-tuning hyperparameters and selecting features improved the Random Forest model's reliability and effectiveness.

This work improves predictive modeling for respiratory illnesses in young people and addresses a gap in incorporating explainability into ML algorithms. This article is organized as follows: The Literature Review examines previous studies, the Materials and Methods section details the information and techniques used, the Results Analysis presents the experimental results, and the Conclusion summarizes the main findings and offers recommendations for future research. The study has a clear final objective. The study's ultimate goal is to enhance pediatric healthcare through AI and ML.

2. Literature Review

Recent studies that have improved the prediction and classification of pediatric respiratory diseases using ML have demonstrated the potential of AI in therapeutic contexts. The two main objectives of survival prediction research are enhancing patient outcomes and supporting clinical decision-making. The techniques, conclusions, and contributions of significant studies are examined in this section.

Roshan Kumar et al. [1] predicted the survival rates of juvenile respiratory illnesses using data from Morocco's Children's Hospital in Rabat. They used the KBest feature selection method to identify 15 significant qualities. AdaBoost, CatBoost, XGBoost, Logistic Regression, and Decision Trees were the experimental classifiers used. The Random Forest with stacking models achieved the highest accuracy of 96%.

Johayra Prithula et al. [4] used data from Zhejiang University's Children's Hospital to anticipate critical care unit mortality in 2024. After reducing 105 characteristics to 16 using different ML models were evaluated. Although CatBoost had an Area Under the Curve (AUC) of 72.22%, a subdivision strategy improved accuracy to 89.32% and AUC to 85.2%. Neel Shah (2022) compared ML and AI in pediatric critical care [25]. The study not only highlighted the importance of ML in clinical decision assistance but also identified challenges with model interpretability and data accessibility. Incorporating these methods into clinical practice involves carefully considering both technical capabilities and real-world requirements.

Sidney Lee et al. 2019 [26] utilized data from UCSF Medical Center and ML techniques to forecast severe sepsis in pediatric patients. Their model, employing 4-fold cross-validation, demonstrated an AUC-ROC of 0.916 at the onset of the condition and 0.718 four hours prior to onset. Nicole et al. [27] conducted a

wide scan of the implementation of ML in chronic childhood respiratory diseases. The review, which had 25 studies from 2013 to 2021, found that 80% of the studies focused on asthma, rather than on other diseases like bronchiolitis, wheezing, or Cystic Fibrosis (CF). Several challenges were identified, such as inconsistent reporting, insufficient validation procedures, and poor reproducibility.

Haijun et al. 2014 [28] developed an ML model to predict transfers to the PICU, using data obtained from Cincinnati Children's Hospital. A Logistic Regression model with 29 variables had a sensitivity of 0.849, a specificity of 0.859, and an AUC of 0.912. The model presented demonstrates the potential for using ML to identify early indicators and plan for children's resources. Spathis et al. [22] used ML approaches to forecast chronic respiratory diseases. The data needed for training ML models was obtained from a clinic in Thessaloniki, Greece, between 2014 and 2015. An accuracy of 97.7% was achieved using the Random Forest classifier. Aykanat et al. [23] applied ML to classify lung diseases. Six datasets were provided for the training of the model from Yıldırım Beyazıt University, Yıldırım Beyazıt Education and Research Hospital, and Ankara University. Allgaier et al. [23] used ML to predict pathogens upon admission by examining clinical characteristics of respiratory infections and pneumonia in hospitalized children. Data from hospitalized patients at National Taiwan University Hospital from 2010 to 2018 was used to train the model. XGBoost achieved a 95% accuracy rate. SHAP was used to assess the importance of each feature. Chang et al. [29] used ML to forecast asthma in children using data from the National Survey of Children's Health in the United States. Random forest surpassed linear regression, Decision Trees, KNN, and Naive Bayes with an accuracy of 90.9%.

These findings highlight the growing importance of ML in pediatric medicine, particularly in the treatment and prognosis of severe respiratory disorders. They show a distinct trend toward increasing predictability, interpretability, and transparency—all of which are critical for the effective use of AI in clinical practice. Nonetheless, there are still issues with data quality, model validation, and the creation of reliable, broadly applicable models. To improve patient outcomes via better informed and timely decision-making, further research is required to improve the application of ML in pediatric healthcare and refine current approaches.

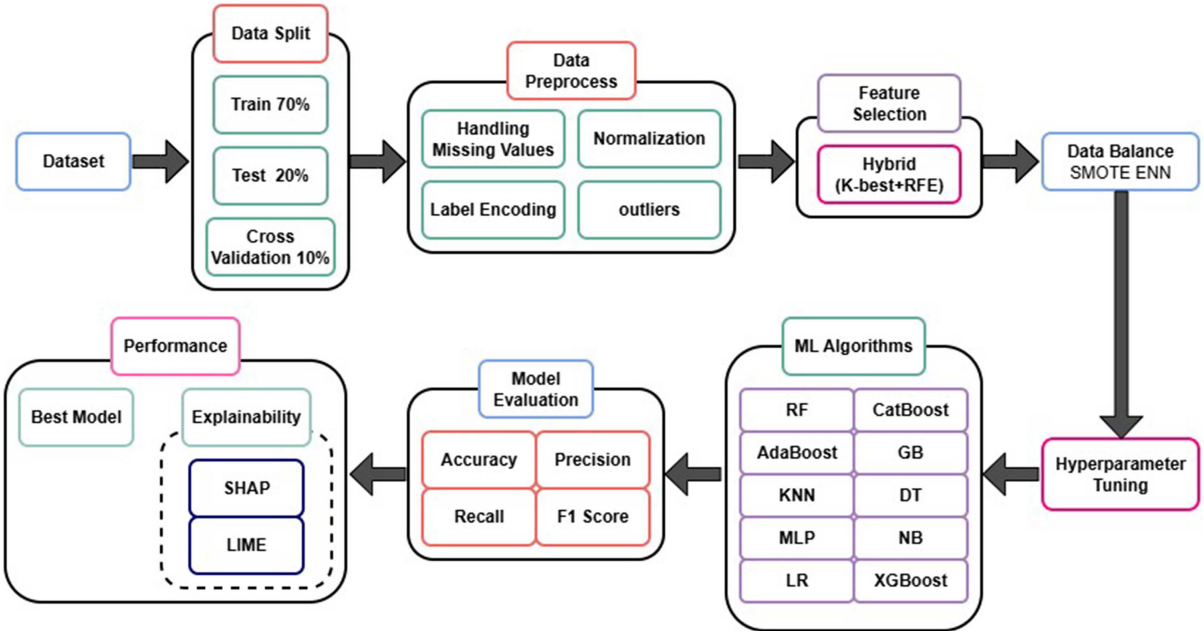
3. Research Methodology

Numerous earlier studies employed a variety of ML models and algorithms to identify suitable answers to issues pertaining to the healthcare sector [30]. However, in addition to the factors that come before the model, like the proper arrangement of the data and the determination of weights according to the nature of each model, the options for each model, which is used as the training set, the cross-validation, and the division of the ratio between data training and testing, the choice of each classifier for ML primarily depends on the experiment in order to evaluate the percentage of accuracy and find the error rate for the metrics. In our thesis, we investigated a few models, evaluated their outcomes based on prior research and a survey of the literature, and then chose the model that best fit the data. Our objective is to use a medical dataset to increase the model's accuracy in forecasting the survivability rate of pediatric respiratory disease. The proposed pediatric respiratory in Figure 1 prediction method involves data preprocessing, hybrid feature selection, sampling, hyperparameter tuning, and model interpretation. The KBest feature selection method and the Recursive Feature Elimination (RFE) method were used to select the top 31 qualities that support critical computations. Throughout the preliminary stage of data processing, strong cleaning techniques and tedious filtering are implemented. Ten classification algorithms are employed alongside hyperparameter tuning to accurately predict the survivability rate of pediatric respiratory disease.

3.1. Description of dataset

Mendeley provided the dataset [31] used in this work, which was utilized in the Children's Hospital of Rabat, Morocco, to treat pediatric respiratory infections. Ninety characteristics from 801 patient records are included, including 65 categorical and 26

Figure 1
The operational procedure of the proposed approach



continuous parameters that are connected to the severity of the illness in children younger than 6.

3.2. Data preprocessing and statistical analysis

Null value deletion, parameter encoding, data standardization, and data balance are some of the processes involved in data preparation. The mean, mode, and median values are typically used in place of the null in several approaches for removing null data. The model predicts clinical development based on factors like weight, age, length of stay, oxygen saturation (SaO2) upon admission, respiratory rate, and heart rate. The high incidence in infants between the ages of 3 and 9 months, as indicated by age analysis in Figure 2, supports its value in predictive modeling.

As seen in Figure 3, heart rate has demonstrated a considerable discriminative potential in predicting clinical outcomes, with a notably higher median value in deceased patients. This implies that a high heart rate is a vital physiological indicator linked to a bad prognosis.

However, as seen in Figure 4, the length of stay was lower for patients who did not survive. This demonstrates its significance as an early indication of disease severity and shows rapid clinical deterioration.

3.2.1. Dealing with missing values

The imputer was employed to deal with missing values. An imputer is used to fill in the missing values in a dataset. Instead of using null values to fill in the numerical values, the median has been utilized to determine the survival rate in pediatric

Figure 2
Age-related risk of infection for pediatric respiratory diseases

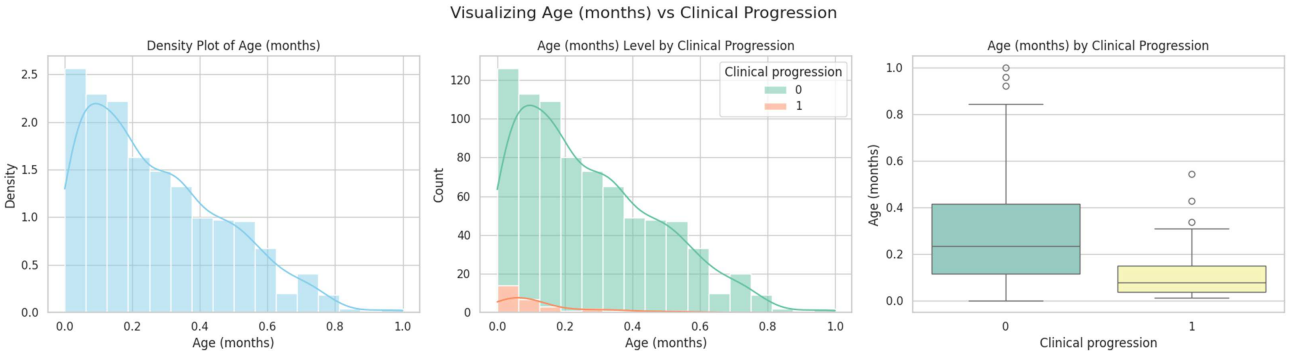


Figure 3
Heart rate-related risk of infection for pediatric respiratory diseases

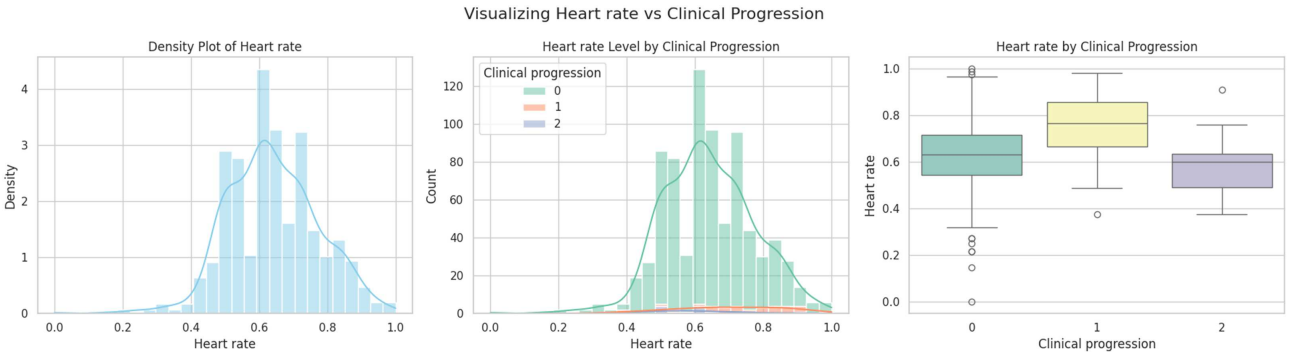
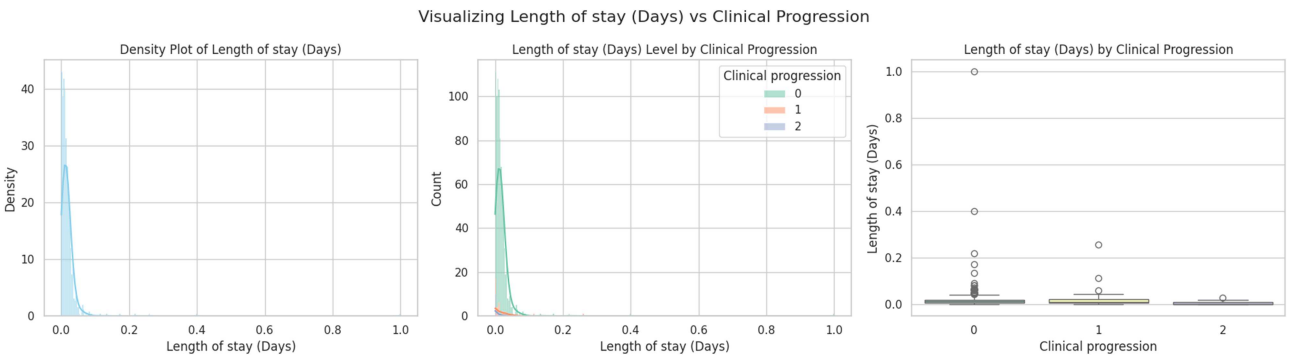


Figure 4
Length of stay-related risk of infection for pediatric respiratory diseases



respiratory disorders. In addition, the category column's missing values have been filled in by mode. Additionally, no garbage values or duplicates were found. Here is a description of the assessed imputers:

1) Median Imputation

In our dataset, we have replaced the missing values with the median. Equation 1 is the formula for calculating the missing value:

$$x_{\{\{(new)\}\}} = \{(median)\}\{(x_1, x_2, \ldots, x_n)\} \quad (1)$$

Median is more robust to outliers than the mean value.

2) Mode Imputation (for categorical data)

The value that occurred most often in the dataset was used to fill in the missing values for categorical data. In Equation 2, the formula of counting mode imputation is given:

$$x_{\{\{(new)\}\}} = \{(mode)\}\{(x)\} \quad (2)$$

3) Data Analysis

Before analyzing data, three columns such as patient ID, admission date, and gender date have been dropped to make the prediction more accurate. Numerous statistical measures, including the median, mode, interquartile range (IQR), and limits, are employed in this work. Here is a description of the measures:

$$Med(X) = \begin{cases} \left(\frac{(X[n/2]) + X[n/2+1]}{2}\right) & \text{if } n \text{ is even,} \\ X[(n+1)/2] & \text{if } n \text{ is odd} \end{cases} \quad (3)$$

a. Median: The middle value in an ordered dataset. If there are odd numbers, it is the center value; if there are even numbers, it is the mean of the two central values. Equation 3 shows the formula for calculating the median:

b. Mode: The value that appears most frequently in a dataset. Equation 4 is the formula of calculating mode.

$$Mode(Z) = L + \left(\frac{(f_m + f_{m-1})}{(f_m - f_{m-1}) + (f_m - f_{m-1})}\right) \times c \quad (4)$$

c. Interquartile Range (IQR): Equation 5 is the formula for IQR, which evaluates the dissemination of data.

$$IQR = Q3 - Q1 \quad (5)$$

here,
Q1 is the first quartile = 25% of the data
Q3 is the third quartile = 75% of the data

d. Lower Bound: Utilized in the detection of outliers. In Equation 6, the lower bound formula has been mentioned:

$$Lower\ Bound = Q1 - (1.5 \times IQR) \quad (6)$$

e. Upper Bound: Utilized in the detection of outliers. The upper bound formula is shown in Equation 7.

$$Upper\ Bound = Q3 + (1.5 \times IQR) \quad (7)$$

The top 31 clinical and demographic characteristics from a dataset are presented in Table 1 along with a data quality summary. Weight, height, heart rate, cyanosis, dehydration signs, oxygen saturation, and other clinical symptoms or treatments (e.g., pleural effusion, dehydration, cephalosporin use) are among the numerical and categorical variables that are included. While categorical features mostly display missing counts, numerical

Table 1
Data quality summary for top 31 features

Feature	Column type	Median	Mode	IQR	Lower bound	Upper bound	Missing values count
Weight (kg)	Numerical	11	10	5.50	0.25	22.25	4
Height (cm)	Numerical	82	70	23	35.5	127.50	32
Heart rate	Numerical	125	120	31.25	63.125	188.1250	13
Procalcitonin	Numerical	0.14	0.05	0.56	–0.78	1.46	17
Patient transferred to intensive care unit	Categorical	NaN	No	NaN	NaN	NaN	0
Detection of DNA/RNA (True-Science Respifinder Pathogen Identification Panel) Allele 1	Categorical	NaN	N/A	NaN	NaN	NaN	245
Detection of DNA/RNA (True-Science Respifinder Pathogen Identification Panel) Allele 2	Categorical	NaN	N/A	NaN	NaN	NaN	49
Antibiotherapy during hospitalization	Categorical	NaN	No	NaN	NaN	NaN	0
Cyanosis	Categorical	NaN	No	NaN	NaN	NaN	0
Dehydration signs	Categorical	NaN	No	NaN	NaN	NaN	0
Paleness	Categorical	NaN	No	NaN	NaN	NaN	0

(*Continued*)

Table 1
(*Continued*)

Feature	Column type	Median	Mode	IQR	Lower bound	Upper bound	Missing values count
Disorders of consciousness	Categorical	NaN	No	NaN	NaN	NaN	1
Gentamicin	Categorical	NaN	No	NaN	NaN	NaN	0
Respiratory rate	Numerical	56	60	14	27	83	0
Hypoventilation	Categorical	No	No	NaN	NaN	NaN	0
Main diagnostic	Categorical	NaN	Asthma attack	NaN	NaN	NaN	0
Nasopharyngeal aspiration	Categorical	NaN	Yes	NaN	NaN	NaN	0
Rhonchi	Categorical	NaN	Yes	NaN	NaN	NaN	0
Age	Numerical	18	3	23	–25.50	66.5	0
Axillary temperature (°C)	Numerical	37.50	37	1.30	35.05	40.25	3
C-reactive protein	Numerical	12	1	33.00	–46.50	85.50	0
Chest X-ray finding	Categorical	NaN	Normal	NaN	NaN	NaN	0
Cephalosporin	Categorical	NaN	No	NaN	NaN	NaN	0
Diagnosis at admission	Categorical	NaN	Asthma attack	NaN	NaN	NaN	0
Duration of pain before consultation (days)	Numerical	2.00	2.00	2	–1.00	7.00	2
Length of stay (days)	Numerical	4	3	5	–5.50	14.50	0
Confirmation of dehydration	Categorical	NaN	5%	NaN	NaN	NaN	791
HIV test results	Categorical	NaN	Negative	NaN	NaN	NaN	796
Nasopharyngeal aspirate culture	Categorical	NaN	Negative	NaN	NaN	NaN	6
Number of persons living in house	Numerical	5	4	2	1	9	2
Oxygen saturation (SaO2) at admission	Numerical	96	98	5	85.50	105.50	34

features report important statistics such as median, mode, IQR, lower and upper bounds, and missing values. This synopsis aids in comprehending the dataset's primary tendencies, variability, and quality.

Along with the above data quality summary, the regression and descriptive statistics summary is also important for understanding the data. So, Table 2 has been integrated. In statistical and ML analyses, parameters such as beta (β) coefficients, confidence intervals, correlations, *p*-values, and group means are crucial for interpreting the strength, direction, and reliability of relationships between variables. The β coefficient represents the effect size of an independent variable on the dependent variable, helping determine how much the outcome changes when the predictor changes by one unit. Its associated confidence interval, β-CI-lower and β-CI-upper, indicates the range within which the true population effect is

Table 2
Regression and descriptive statistics summary

Feature	β	β-CI-lower	β-CI-upper	*Y*-correlation	*p*	Y_1-mean	Y_0-mean	All mean	All Std
Weight (kg)	–0.1721	–0.2994	–0.0448	–0.0935	0.0081	nan	nan	0.3864	0.1737
Height (cm)	–0.1077	–0.2407	0.0253	–0.0562	0.1123	nan	nan	0.564	0.1667
Heart rate	0.1357	–0.0309	0.3023	0.0565	0.1101	nan	nan	0.6401	0.1331
Procalcitonin	0.79	0.3019	1.2781	0.1117	0.0015	nan	nan	0.008	0.0452
Patient transferred to intensive care unit	0.4068	0.3305	0.483	0.3476	0	nan	nan	0.0811	0.2732
Rhonchi	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Yes	NaN
Detection of DNA/RNA (True-Science Respifinder Pathogen Identification Panel) Allele 1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N/A	NaN

(*Continued*)

Table 2
(*Continued*)

Feature	β	β-CI-lower	β-CI-upper	*Y*-correlation	*p*	Y_1-mean	Y_0-mean	All mean	All Std
Detection of DNA/RNA (True-Science Respifinder Pathogen Identification Panel) Allele 2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N/A	NaN
Cyanosis	0.1987	0.1219	0.2756	0.1767	0	nan	nan	0.0886	0.2844
Dehydration signs	0.2649	0.0831	0.4466	0.1007	0.0043	nan	nan	0.015	0.1216
Paleness	0.1038	0.0444	0.1632	0.1205	0.0006	nan	nan	0.1648	0.3712
Disorders of consciousness	0.2388	0.1069	0.3706	0.1248	0.0004	nan	nan	0.0287	0.1671
Gentamicin	0.2	0.1014	0.2986	0.1395	0.0001	nan	nan	0.0524	0.223
Respiratory rate	0.028	–0.2335	0.2894	0.0074	0.8338	nan	nan	0.2081	0.0849
Hypoventilation	0.1954	0.0502	0.3406	0.093	0.0084	nan	nan	0.0237	0.1523
Main diagnostic	0.0061	0.0037	0.0085	0.1744	0	nan	nan	9.3421	9.1238
Nasopharyngeal aspiration	0.004	–0.0017	0.0097	0.0487	0.1685	nan	nan	6.8052	3.8765
HIV test results	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Negative	NaN
Age (months)	–0.0438	–0.153	0.0653	–0.0279	0.4306	nan	nan	0.2693	0.2034
Axillary temperature (°C)	–0.0367	–0.2131	0.1397	–0.0144	0.6832	nan	nan	0.3932	0.1259
C-reactive protein	0.0606l	–0.0862	0.2074	0.0286	0.4181	nan	nan	0.1184	0.1512
Chest X-ray finding	–0.0164	–0.0434	0.0105	–0.0423	0.2317	nan	nan	1.6192	0.8238
Cephalosporin	0.0443	–0.004	0.0925	0.0636	0.072	nan	nan	0.3021	0.4595
Diagnosis at admission	0.005	0.0018	0.0082	0.108	0.0022	nan	nan	8.9563	6.9348
Duration of pain before consultation (days)	0.044	0.005	0.082	0.086	0.026	6.310	3.839	3.929	5.366
Length of stay (days)	–0.1139	–0.6484	0.4206	–0.0148	0.6759	nan	nan	0.0182	0.0415
Confirmation of dehydration	NaN	NaN	NaN	NaN	NaN	NaN	NaN	<5%	NaN
Antibiotherapy during hospitalization	NaN	NaN	NaN	NaN	NaN	NaN	NaN	No	NaN
Nasopharyngeal aspirate culture	–0.021	–0.0458	0.0038	–0.0587	0.0972	nan	nan	3.7828	0.8932
Number of persons living in house	–0.1056	–0.2583	0.047	–0.048	0.1747	nan	nan	0.26	0.1453
Oxygen saturation (SaO2) at admission	–0.1021	–0.27	0.0658	–0.0422	0.2328	nan	nan	0.8414	0.1322

expected to lie, providing a measure of estimate precision. The *p*-value (*p*) tests the statistical significance of this relationship—that is, whether the observed effect is likely due to chance. The Y-correlation reflects the linear relationship between the predictor and the response variable, giving insight into their direct association. Meanwhile, Y1-mean and Y0-mean show the average values of the dependent variable for two different groups, for example, treatment vs control, revealing the difference in outcomes between them. Finally, the all mean and all standard deviation (Std) summarize the overall distribution of the outcome variable, helping to understand its central tendency and variability. Together, these criteria enable researchers to evaluate not just the strength and direction of correlations but also their dependability, precision, and real-world interpretability.

f. β (Beta Coefficient): It shows the amount of change in the dependent variable (Y) resulting from a one-unit change in the independent variable (X), while keeping all other variables constant. It assesses the strength and direction of the link between X and Y. A simple linear regression is calculated using the formula in Equation 8:

$$\beta = \mathrm{Cov}(\mathrm{X}, \mathrm{Y}) / \mathrm{Var}(\mathrm{X}) \tag{8}$$

where Cov(X, Y) is the covariance between X and Y and Var(X) is the variance of X.

g. Lower and Upper Bounds of the β-Confidence: This interval defines the range within which the true population value of β is expected to be found with a certain level of confidence, often 95%. This range helps evaluate the precision of the estimated coefficient. The formula in Equation 9 is for the confidence interval:

$$\mathrm{CI} = \beta \pm \mathrm{t}_{\mathrm{a}/_2} \times \mathrm{SE}\beta \tag{9}$$

where $t_{a/2}$ is the critical value from the *t*-distribution and SEβ is the standard error of β.

h. Y-correlation: It is also known as Pearson's correlation coefficient, which measures the strength and direction of the linear relationship between two continuous variables, X and Y. It is calculated in Equation 10, as

$$r = \Sigma[X_i - \overline{X}Y_i - \overline{Y}]/\sqrt{[\Sigma X_i - \overline{X}^2 \times \Sigma Y_i - \overline{Y}^2]} \tag{10}$$

where $\bar{X}$ and $\bar{Y}$ are the means of X and Y, respectively. The value of r ranges from −1 to +1, indicating the strength of negative or positive correlation.

i. p-value: p indicates the probability of obtaining a test statistic at least as extreme as the one observed, assuming the null hypothesis is correct. It helps determine the statistical significance of the result. A smaller p-value typically < 0.05 indicates substantial evidence against the null hypothesis. This value is obtained from the test statistic formula, which is in Equation 11:

$$t = \beta / SE(\beta) \tag{11}$$

where the probability $p = P(|T| > t)$ is calculated using the t-distribution.

j. Y_1-mean: It represents the average value of the dependent variable for group 1, often called the treatment or exposed group. It is determined by Equation 12:

$$\bar{Y}_1 = \Sigma Y_1 / n_1 \tag{12}$$

where ΣY_1 is the sum of Y values in group 1 and n_1 is the number of observations in that group.

k. Y_0-mean:

It signifies the average of the dependent variable for group 0, which usually represents the control or unexposed group. It is computed as Equation 13:

$$\bar{Y}_0 = \Sigma Y_0 / n_0 \tag{13}$$

where ΣY_0 is the total of Y values in group 0 and n_0 is the number of observations in that group.

l. All mean: This metric indicates the overall average of the dependent variable across all observations, without considering grouping. It provides a general measure of central tendency and is calculated by Equation 14:

$$\bar{Y} = \Sigma Y_i / N \tag{14}$$

where N represents the total number of observations.

m. Standard Deviation: This metric evaluates the dependent variable's dispersion or variability around the mean. It indicates how far the data points differ from the average. The formula is as Equation (15):

$$s = \sqrt{[\Sigma (Y_i - X)^2 / (N - 1)]} \tag{15}$$

where Y_i represents each observation, ɲ is the overall mean, and N is the total number of observations.

3.3. Normalization

To prevent errors, the data must be normalized because the attributes in the obtained data may have different scales. Normalization has been used to rescale the data within a specified range in order to prevent errors. The process of normalization involves scaling data into a range, usually [0, 1]. It guarantees that each feature makes an equal contribution to model training. Another name for it is min–max scaling. In equation 8, the normalization formula is mentioned in Equation 16:

$$x_{norm} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \tag{16}$$

3.4. Removing outlier

Outliers have been evaluated from the dataset after min–max scaling. An outlier is a data point that significantly deviates from the other values in a dataset. The IQR is a widely used technique for eliminating outliers.

Any value is an outlier in Equation 17:

$$Either\ x > \ Upper\ Bound\ or\ x < \ Lower\ Bound \tag{17}$$

The dataset's size drops to -(560,87) once the outlier is eliminated, indicating that the majority of the individual values were outside the lower and upper whisker range. Categorical values were then encoded using label encoding. Table 2 contains the specifics of the data encoding for the top features.

3.5. Label encoding

Label encoding is a widely used method for managing categorical data. This approach assigns a unique number to each category based on the label order. The encoded values can be found using Equations (18) and (19):

$$S = \frac{1}{1 + \exp(-\frac{n - mdl}{a})} \tag{18}$$

$$\hat{x}^k = prior \times (1 - s) + s \times \frac{n^+}{n} \tag{19}$$

Here, the smoothing parameter "an" is used to assess the effectiveness of regularization, and "mdl" is the minimal number of data samples required at a leaf node. It makes sense to assume that the mdl and values range from 1 to 100. Old values are replaced with new ones unique to the training dataset, as described in [32].

The normalized ranges of the dataset's numerical and categorical features are compiled in Table 3. Weight, height, heart rate, creatinine, age, and oxygen saturation are examples of numerical variables whose values were scaled from 0.0 to 1.0 according to their lowest and highest values. The binary representation of categorical data (such as cyanosis, dehydration, confirmation of dehydration, antibiotherapy during hospitalization, and diagnostic results) is "No–Yes." All things considered, the table emphasizes the preprocessing stage of encoding and normalization used to guarantee that every feature is on the same scale for additional modeling or analysis.

Table 3
Data encoding information for top features

Feature	Column type	Min	Max	Range after normalization
Weight (kg)	Numerical	2.9	24	0.0–1.0
Height (cm)	Numerical	32	120	0.0–1.0
Heart rate	Numerical	14	190	0.0–1.0
Respiratory rate	Numerical	21	195	0.0–1.0
Procalcitonin	Numerical	0.04	370.38	0.0–1.0
Patient transferred to intensive care unit	Categorical	No	Yes	No–Yes
Cephalosporin	Categorical	No	Yes	No–Yes
Cyanosis	Categorical	No	Yes	No–Yes
Dehydration signs	Categorical	No	Yes	No–Yes
Disorders of consciousness	Categorical	No	Yes	No–Yes
Gentamicin	Categorical	No	Yes	No–Yes
Hypoventilation	Categorical	No	Yes	No–Yes
Main diagnostic	Categorical	No	Yes	No–Yes
Nasopharyngeal aspiration	Categorical	No	Yes	No–Yes
Detection of DNA/RNA (TrueScience Respifinder Pathogen Identification Panel) Allele 1	Categorical	No	Yes	No–Yes
Detection of DNA/RNA (TrueScience Respifinder Pathogen Identification Panel) Allele 2	Categorical	No	Yes	No–Yes
Paleness	Categorical	No	Yes	No–Yes
Age (months)	Numerical	1	78	0.0–1.0
Axillary temperature (°C)	Numerical	35	42	0.0–1.0
C-reactive protein	Numerical	0	220	0.0–1.0
Chest X-ray finding	Categorical	No	Yes	No–Yes
Diagnosis at admission	Categorical	No	Yes	No–Yes
Rhonchi	Categorical	No	Yes	No–Yes
Length of stay (days)	Numerical	0	333	0.0–1.0
Duration of pain before consultation (days)	Numerical	1	60	1.0–60.0
Confirmation of dehydration	Categorical	No	Yes	No–Yes
Antibiotherapy during hospitalization	Categorical	No	Yes	No–Yes
Number of persons living in house	Numerical	1	18	0.0–1.0
Oxygen saturation (SaO2) at admission	Numerical	68	100	0.0–1.0
HIV test results	Categorical	No	Yes	No–Yes

3.6. Data splitting

To make sure that ML algorithms are generalizable, data partitioning is essential. It is a two-step data splitting procedure that uses stratification to maintain the class distribution while preparing the dataset for testing, validation (cross-validation), and training. The "Clinical progression" column was removed from the dataset and assigned to the target column in order to first separate the features and target variable. The class distribution was then maintained throughout splits by performing a stratified split to reserve 20% of the data as a hold-out test set. Model development uses the remaining 80%, where Stratified *K*-Fold Cross-Validation [33] with 8 splits has been used. In each fold, this method maintains class balance by allocating roughly 70% of the total data for training and 10% for validation. For iterative model training and assessment, training and validation sets inside the cross-validation loop have been generated.

3.7. Feature selection

To lower dimensionality and keep the most valuable predictors for the classification task, a hybrid feature selection technique has been implemented in this pipeline. To choose the best features from the training set, two tried-and-true methods have been implemented after dividing the dataset into training, validation, and testing sets using stratified sampling to maintain class distribution.

To select the top 20 features, first, the SelectKBest feature selection method has been implemented, and then again, RFE has been implemented to select the top 20 features. Finally, both of the 20 features have been gone through the feature union procedure.

1) SelectKBest feature selection using the F-test for ANOVA: This approach selects features according to the top k highest scores derived from a given scoring function, such as mutual information, the chi-squared test, or the ANOVA F-test. Based

on their statistical importance in relation to the target variable, this method is particularly useful for maintaining the most relevant traits.

With this approach, features are ranked according to how statistically they relate to the target variable. An F-score is calculated to evaluate each attribute independently by comparing the variance between classes to the variance within classes. The top $k = 20$ characteristics with the highest F-scores are chosen because they are believed to have the strongest linear relationship with the target.

SelectKBest employs the ANOVA F-test as a filter method to assess the significance of features using statistical methods. It uses the ANOVA F-test, a univariate metric that evaluates the significance of each feature in relation to the target variable. The goal is to select the "k" best characteristics, where "k" is a parameter specified by the user [34]. The ANOVA F-statistic between each feature and the target variable is calculated using the ANOVA F-test function. This statistic, which measures the degree of linear dependency between the feature and the target, can be used to identify the features most likely to be relevant for classification. Equation 20 is the formula for the F-statistic:

$$F = SSB/(k-1)/SSW(n-k) \tag{20}$$

where k is the number of classes, n is the total number of data points, and the sum of squares between (SSB) and within (SSW) reflects the variance between and within each group [34]. Combining the ANOVA F-test with the feature selection algorithm SelectKBest results in SelectKBest with the ANOVA F-test. This method's main goal is to assess each feature's importance in relation to a target variable. It functions as a filter approach, which means that the learning algorithm is not used to rank features according to statistical measurements. In this case, a statistical test that determines if the means of many groups are equal is the ANOVA F-test. It aids in measuring the correlation between each feature and the target variable in feature selection [35].

2) Random Forest with Recursive Feature Elimination (RFE): RFE is a wrapper-based technique that uses model performance to recursively remove the least significant features. The estimator in this case is a Random Forest classifier, which rates features according to how well they lower impurity in Decision Trees. Up until the top 20 most crucial features are chosen, the RFE procedure iteratively eliminates features and refits the model.

RFE is a feature selection technique that removes the weakest features by fitting a model. The best ranking set of features is chosen by combining cross-validation and RFE to determine the ideal feature count. The caret package can be used to implement the RFE procedure in the R tool. First, the RFE algorithm's control function needs to be defined. Here is a statement of the Random Forest selection function over the rfFuncs option in the rfeControl function: function = rfFuncs, method = "cv", number = 10; control< - rfeControl.

The RFE algorithm is then put into practice as follows. Sizes = 1:10, rfeControl = control, training_data[1:22], training_data, rfe.train < - rfe. The process was carried out using predictor variables and a newly added target variable after the original target variables in the dataset were eliminated. Following the RFE algorithm's implementation, the outcome was displayed, yielding a variable significance chart.

After using both methods, the union of the two feature lists had been chosen and produced a hybrid feature set. This guarantees that any feature that each approach deems significant is kept. By balancing the advantages of wrapper techniques (like RFE, which takes feature interactions into account) with filter techniques (like SelectKBest, which is quick and simple), this method makes feature selection more thorough and reliable. In order to create a cleaner, more targeted dataset for model training and evaluation, you then shrunk both your training and validation datasets to contain only these chosen characteristics.

The top 31 variables by the number of features that are important for predicting outcome are shown in Table 4 and ranked in order of importance. "Patient transferred to intensive care unit" (score 128.47) is the most significant aspect, followed by "cyanosis" (27.76) and "disorders of consciousness" (26.39). High-ranked clinical signs are antibiotics (cephalosporin, gentamicin), signs of dehydration, procalcitonin, and hypoventilation. Conversely, general physiological and demographic characteristics, including temperature, duration of stay, and respiration rate, get lower scores. This ranking shows that, by far, severe clinical symptoms and critical care indicators are the biggest contributors to predictive modeling.

Table 4
Results of top 31 feature scoring

No.	Top feature	Score	Descriptions
01	Patient transferred to intensive care unit	128.465208	If a patient got transferred to ICU or not
02	Detection of DNA/RNA (TrueScience Respifinder Pathogen Identification Panel) Allele 1	0.454706	Identifies the presence of a specific respiratory pathogen via molecular testing
03	Disorders of consciousness	26.393402	Surveillance on semi-conscious or unconscious patients
04	Cyanosis	27.758907	Bluish skin discoloration caused by low oxygen in blood
05	Procalcitonin	7.641946	Blood biomarker for bacterial specificity

(*Continued*)

Table 4
(*Continued*)

No.	Top feature	Score	Descriptions
06	Hypoventilation	19.991707	Abnormally slow breathing
07	Gentamicin	12.366395	Antibiotic used to treat bacterial infections
08	Weight (kg)	12.457664	Patient's weight in kilograms
09	Nasopharyngeal aspiration	13.037221	Suctioning technique to clear the mouth and nose of mucus or fluids
10	Dehydration signs	13.382234	Key symptoms of dehydration (like dry mouth, sunken eyes, low urine output)
11	Heart rate	9.450807	Number of heartbeats per min
12	Chest X-ray finding	6.258920	Identifies abnormalities in chest X-rays
13	Cephalosporin	9.028784	Antibiotics used to manage bacterial infections
14	Main diagnostic	6.284291	The primary medical condition a patient is being treated for
15	Detection of DNA/RNA (TrueScience Respifinder Pathogen Identification Panel) Allele 2	0.328437	Detects an additional pathogen or co-infection genetically
16	Paleness	13.596838	Decreased blood supply to the skin
17	Rhonchi	5.883983	Abnormal lung sound indicating airway obstruction or mucus presence
18	Height (cm)	7.264458	Patient's height in centimeters
19	HIV test results	13.901820	Indicates patient's HIV infection status
20	Age (months)	6.354414	Patient's age in months
21	Duration of pain before consultation (days)	1.548219	Number of days patient had pain before seeking care
22	Nasopharyngeal aspirate culture (NPA)	2.673056	A test mucus suctioned from the nose to identify infections
23	Diagnosis at admission	2.313827	Diagnosis given to patient on admission
24	Confirmation of dehydration	6.049290	Clinical verification of fluid loss/dehydration status
25	Oxygen saturation (SaO2) at admission	0.157219	Oxygen level in blood at admission
26	Antibiotherapy during hospitalization	6.388288	Indicates whether the patient received antibiotics during hospital stay
27	C-reactive protein	3.330348	Protein made by the liver acts as a nonspecific indicator of inflammation
28	Number of persons living in house	0.605498	Household occupancy count
29	Respiratory rate	0.802916	Breathing rate per min
30	Length of stay (days)	3.349635	How many days has the patient been affected/suffering
31	Axillary temperature (°C)	0.240944	Temperature under the armpit

3.8. Oversampling

A hybrid resampling method SMOTE-ENN has been used to address the class imbalance issue in the training data after feature selection. For balancing, SMOTE (Synthetic Minority Oversampling Technique) generates synthetic instances of the minority class, and ENN (Edited Nearest Neighbors) removes noisy and misclassified instances from both classes. It is advantageous for generalization and reduces the bias toward the majority class, making this method very useful in the context of boosting classification accuracy in imbalanced data. It takes a sample from the minority class and then creates new samples by drawing line segments between the sample and one of its KNN. This improves the model's ability to learn the decision boundaries, particularly if the minority class is not well represented. ENN, on the other hand, is a data cleansing method that removes samples that are unclear or misclassified. It does this by examining the KNN of each data point and removing those with a different class label than their neighbors. This step is critical because it aids in the removal of noisy or ambiguous instances that may confuse the classifier. In medical or critical classification tasks, where missing minority cases can have major repercussions, this combination helps achieve better precision-recall balance, improves sensitivity to the minority class, and improves model generalization.

1) SMOTE: Creating synthetic samples for the minority class is how the first part, known as SMOTE, operates. It accomplishes this by choosing a data point x_i from the minority class and one of its KNN x_{zi}, and then using the following formula of Equation 21 to create a new sample x new x new along the line segment between them:

$$x_{new} = x_i + \lambda \cdot (x_{zi} - x_i) \tag{21}$$

where the random number λ ranges from 0 to 1. By increasing the minority class's representation without repeating already-existing data points, this procedure aids the classifier in better understanding its traits.

2) ENN: After oversampling, the dataset is cleaned by eliminating possibly noisy or confusing samples using ENN. ENN looks at each instance's KNN in the dataset. An instance is deemed misclassified and eliminated if its class label deviates from the majority class of its neighbors. Both majority and minority classes are subjected to this rule-based filtering, which aids in removing examples that are unclear or poorly labeled and may have a detrimental effect on the classifier's performance. SMOTE-ENN works particularly well for enhancing classification results in unbalanced datasets since it produces a more balanced and cleaner dataset when combined.

Clinical progression, the study's objective variable, is separated into three groups: "Non-Severe" (healed and released), "Severe" (death), and "At-Risk" (left against medical recommendation). Inequality was rectified using SMOTE, which generated synthetic minority class data that enhanced F1 score, recall, and classifier performance. As the dataset was an 8-fold CV, within each iteration of the stratified 8-fold cross-validation, SMOTE-ENN was strictly applied to the training folds, and to avoid data leakage, the test folds were kept totally hidden.

3.9. Machine learning models

In order to evaluate the survival rate of juvenile respiratory disorders, scientists employed a range of ML algorithms throughout the current study. To do so, researchers carefully looked at the following ML models—Random Forest, Logistic Regression, AdaBoost, MLP, KNN, Decision Trees, CatBoost, XGBoost, Naïve Bayes, and Gradient Boosting—due to their efficacy in medical diagnosis. GridSearchCV was used to adjust the hyperparameters. The values of the parameters used for each classifier are shown in Table 5.

Table 5
Hyperparameter values for each classifier

Classifiers	Parameters	Values
Random Forest	Class weight	[Balanced]
	Max depth	[20]
	Min samples split	[2]
	n-estimator's	[100]
XGBoost	Learning rate	[0.1]
	Max depth	[3]
	n-estimator's	[200]
Gradient Boosting	Learning rate	[0.1]
	Max depth	[5]
	n-estimator's	[200]
KNN	Metric	[Manhattan]
	N neighbors	[3]
	Weights	[Distance]
Decision Tree	Max depth	[None]
	Min samples split	[5]
Logistic Regression	C	[10]
AdaBoost	Learning rate	[0.1]
	n-estimator's	[100]
MLP	Alpha	[0.001]
	Hidden layer sizes	[(50,50)]
CatBoost	Depth	6
	Iterations	500
	Learning rate	0.1

3.10. Random forest

A well-liked ML method for a variety of classification tasks is Random Forest. An ensemble of classifiers with a tree topology is called a Random Forest. Each input is given the most likely class label based on a unit vote from each tree in the forest. It is a fast and noise-resistant ensemble that successfully identifies nonlinear patterns in the data. It can easily manage both numerical and qualitative data. One of Random Forest's key benefits is that it doesn't overfit even when more trees are introduced to the forest. This method successfully lowers variation and raises the model's overall accuracy by combining the predictions from several Decision Trees. The Random Forest algorithm is based on this principle [36]. The Gini Index, a metric for assessing how well classes are divided, is created by using Equation (22):

$$Gini = 1 - \sum_{J-1}^{n} (p_J)^2 \tag{22}$$

The probability that an object will belong to a specific class or characteristic is represented by the number p_J in Equation (22).

3.11. Environmental setup

After gathering the data, it was split into training and testing sets and then randomly shuffled. To avoid overfitting and data leakage, particular methods were used to construct the testing sets. A precise configuration is necessary for the study to produce the desired outcomes. Table 6 details the configuration and prerequisites for carrying out the investigation.

Table 6
System's specification

Platform	Google Co-Lab
Language	Python
Libraries	Matplotlib, Scikit-Learn, NumPy, Pandas, and Seaborn
RAM	12.7 GB
CPU	i5-1145G7 @ 2.60 GHz
Session Limit	12 h
Disk Space	107.7 GB

3.12. Performance metrics

To assess the categorization of respiratory illnesses, the F1-score, recall, accuracy, and precision were used. In the confusion matrix, the letters T, F, P, and N stand for true, false, positive, and negative results, respectively. TP shows that the detection of a healthy case was done correctly. Equations (23) through (28) define accuracy, precision, F1-score, and recall, where recall measures how well the model detects true positives.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{23}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{24}$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{25}$$

$$\text{F1} = 2 \times \{\frac{Precision\,.\,Recall}{(Precision + Recall)}\} \tag{26}$$

$$\text{Specificity (Macro} - \text{averaged)} = \frac{1}{k}\sum_{i}^{k}\frac{TN}{FP + TN} \tag{27}$$

$$\text{Error Rate (\%)} = (1 - Accuracy) \times 100 \tag{28}$$

4. Result Analysis

In order to predict the survival of juvenile respiratory disorders, the study assessed 10 ML classifiers: Random Forest, Logistic Regression, AdaBoost, MLP, KNN, Decision Trees, CatBoost, XGBoost, Naïve Bayes, and Gradient Boosting that performed better than the others, with an accuracy rate of 99.29%, a precision of 92.68%, a recall rate of 92.55%, and an F1-score of 92.58%. These comprehensive metrics show how reliable it is as a screening tool, reliably and precisely identifying rare occurrences in clinical contexts.

So, from Table 7, it's clear that with an accuracy percentage of 99.15%, Gradient Boosting and CatBoost both showed outstanding performance. These models showed their dependability as alternatives to Random Forest by describing high recall and accuracy scores. With an accuracy of 99.08%, 99.08%, 97.37%, and 96.95%, respectively, XGBoost, MLP, KNN, and Decision Tree fared somewhat better than AdaBoost, which is 91.84%. With an accuracy of 34.78%, Naïve Bayes had the lowest accuracy, indicating that it might not be as helpful for this particular dataset. Despite the model attaining an accuracy of 99.29%, the F1-score of around 93% indicates that the classification of the minority mortality class is more difficult than that of the majority survival class. This prevalent issue in imbalanced clinical datasets underscores the necessity of assessing performance beyond mere accuracy. However, the model has 95% confidence intervals (Bootstrap), along with precision: 92.75 (87.13–97.12), recall: 92.47 (88.20–96.27), F1-score: 92.44 (87.70–96.27), and specificity: 76.17 (65.17–87.62).

Table 7
Results of 10 machine learning models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Specificity	Error rate (%)
Random Forest	99.29	92.68	92.55	92.58	76.49	0.71
XGBoost	99.08	92.97	89.44	91.05	82.42	0.92
Gradient Boosting	99.15	94.51	92.55	93.39	86.97	0.85
KNN	97.37	91.38	67.70	76.68	74.96	2.63
Decision Tree	96.95	95.85	92.55	93.74	90.47	3.05
CatBoost	99.15	93.81	88.82	90.87	85.69	0.85
Logistic Regression	96.10	92.37	80.75	85.63	79.44	3.90
AdaBoost	91.84	93.25	73.29	81.09	80.41	8.16
MLP	99.08	90.38	87.58	88.94	71.30	0.92
Naïve Bayes	34.78	91.80	34.78	48.36	44.58	65.22

Significantly, as Random Forest performed the best in our study, we are adding performance metrics like precision (weighted), recall (weighted), F1-score (weighted), F1-score (macro), specificity (macro), recall (Class 0), recall (Class 1), recall (Class 2), and macro AUC to acknowledge the model's robustness.

The metrics:

Precision (weighted): 0.9268
Recall (weighted): 0.9255
F1-score (weighted): 0.9258
F1-score (macro): 0.5561
Specificity (macro): 0.7649
Recall (Class 0): 0.9605
Recall (Class 1): 0.3333
Recall (Class 2): 0.3333
Macro AUC: 0.9360

The classification accuracy of nine ML models is evaluated by comparing them using a confusion matrix, as shown in Figure 5. The classification accuracy of nine ML models is compared using the confusion matrix, as in Figure 5. The matrix is a pictorial representation of the true positives, false positives, and false negatives. Random Forest, CatBoost, and Gradient Boosting models have the highest accuracy and low misclassification for all classes; predictions are focused on the diagonal elements of these models. Despite slightly higher misclassifications than other models, such as XGBoost or MLP, good performance is still obtained. Decision Tree, KNN, Logistic Regression, and AdaBoost have lower classification performance with misclassifying Class 0 with other classes of the data. Random Forest and CatBoost achieve the highest prediction accuracy in this dataset, and Random Forest is the best in general.

The ROC curves of nine classifiers (Random Forest, XGBoost, Gradient Boosting, AdaBoost, CatBoost, Decision Tree, KNN, Logistic Regression, and MLP) are displayed in Figure 6. The model using Random Forest has consistently high AUC values compared to the other models, indicating that the model has good discriminatory power. AUC values are comparable to those of Random Forest, CatBoost, Gradient Boosting, XGBoost, and Decision Tree and demonstrate high accuracy with

Figure 5
Confusion matrix of different ML classifiers

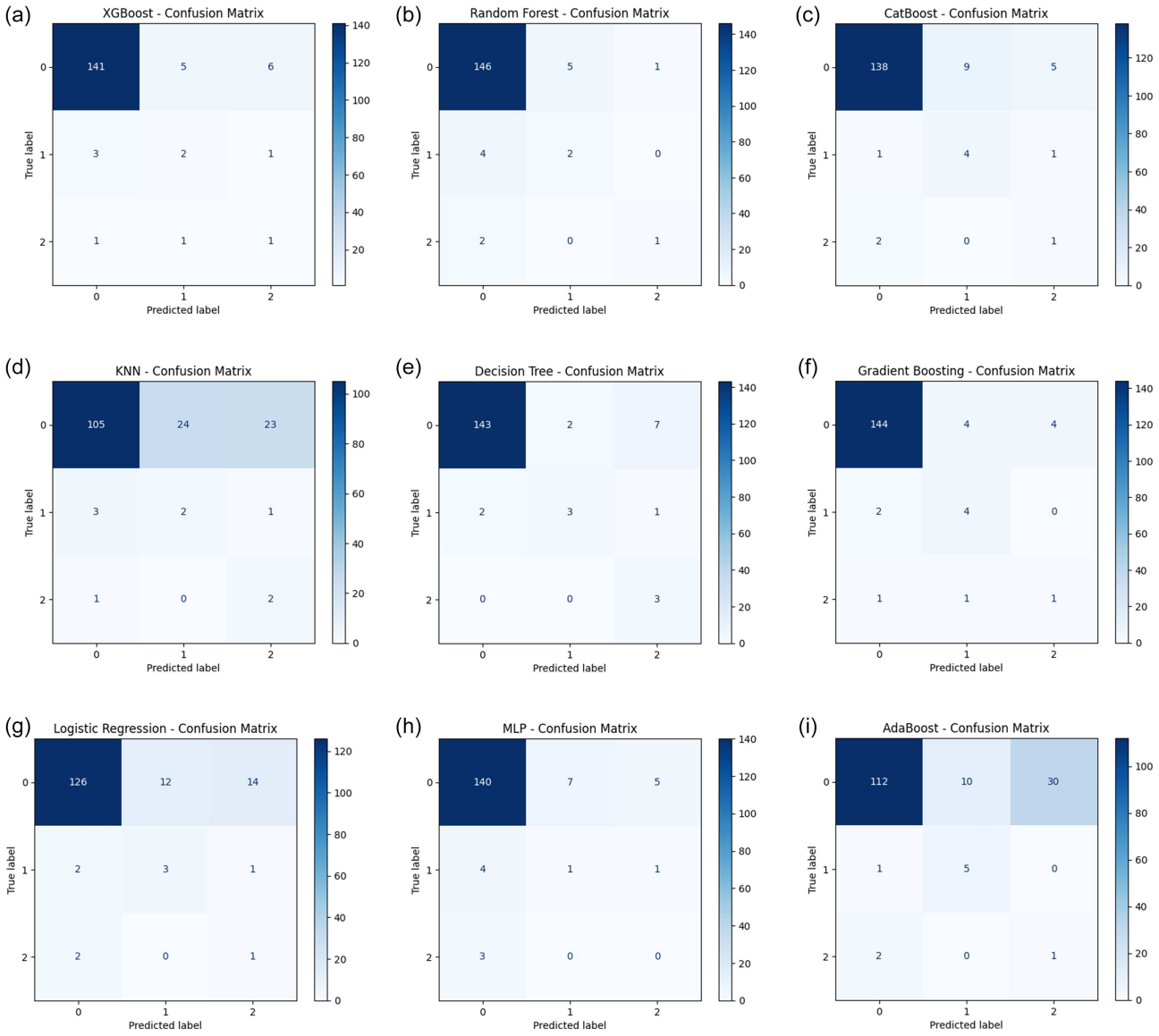
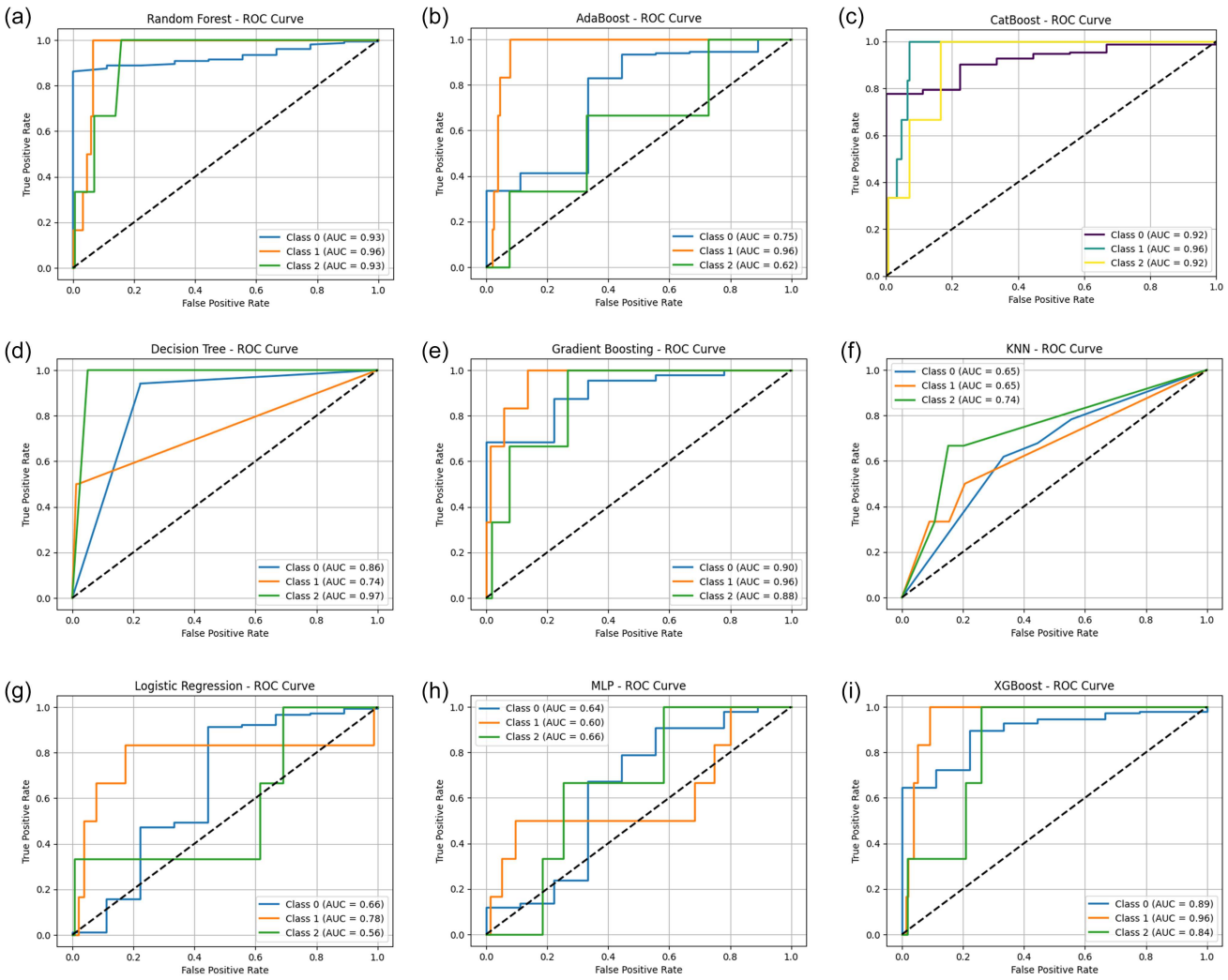


Figure 6
ROC curve comparison



slightly decreased class-to-class stability. The other traditional classifiers, such as Adaboost, KNN, Logistic Regression, and MLP, however, have significantly lower ROC curves accompanied by their corresponding lower AUC. This indicates that these models are not able to perform well in the class separation task. The Random Forest ROC curve is well above the diagonal and near the upper left corner of the plot, further supporting its robustness. The outcome of these findings confirms that the proposed Random Forest model is the most reliable classifier for the clinical prediction problem because it not only achieves the highest accuracy, but it also has the best sensitivity-specificity balance.

5. Model Interpretation

To enhance the model's interpretability, SHAP was applied to the Random Forest classifier. Using Beeswarm plots to show relevance, SHAP assigns a Shapley value to each attribute [37].

The interpretability of an ML model that uses SHAP to predict the "Non-Severe" class is revealed by Figure 7(a–c). Using SHAP and LIME explanation methodologies, these three images show how various characteristics affect an ML model's prediction for the "Non-Severe" class. "Patient transferred to intensive care unit" and "length of stay (days)" are the most significant variables in predicting non-severe cases, according to the SHAP summary bar plot, which offers a global picture of feature relevance. The average absolute contributions of each feature across all samples are summarized in this graphic; however, it is not made clear whether or not they raise or lower the forecast.

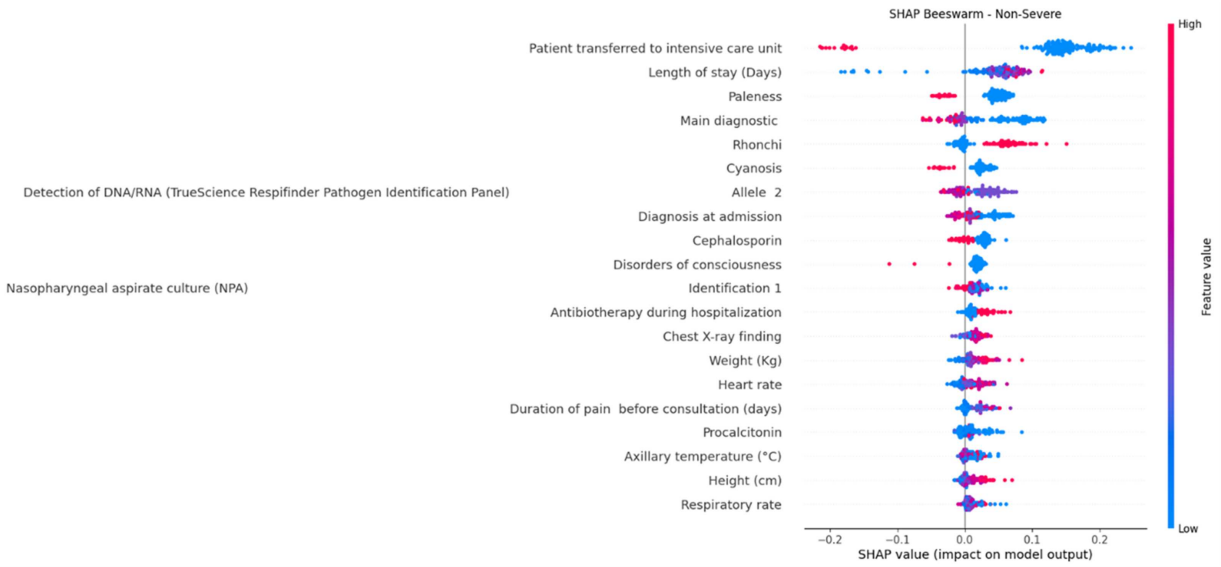
By illustrating the impact of each attribute on individual predictions, the SHAP Beeswarm plot provides additional detail. The direction and size of the impact of each feature are indicated by the SHAP value along the x-axis, and each dot represents a sample. Low patient transferred to intensive care unit (ICU) tends to increase the probability of a non-severe categorization, for example, as the color gradient displays the real feature value (low values in blue, high in red).

Lastly, the LIME Explanation Plot breaks out which particular feature conditions contributed positively (green) or negatively (red) to the classification, concentrating on a single prediction. A non-severe outcome is strongly supported in this instance by factors like having no ICU stay and a restricted X-ray finding. The simplicity and individual-level interpretability of LIME, which provides distinct thresholds for the influence of each variable, are its strongest points.

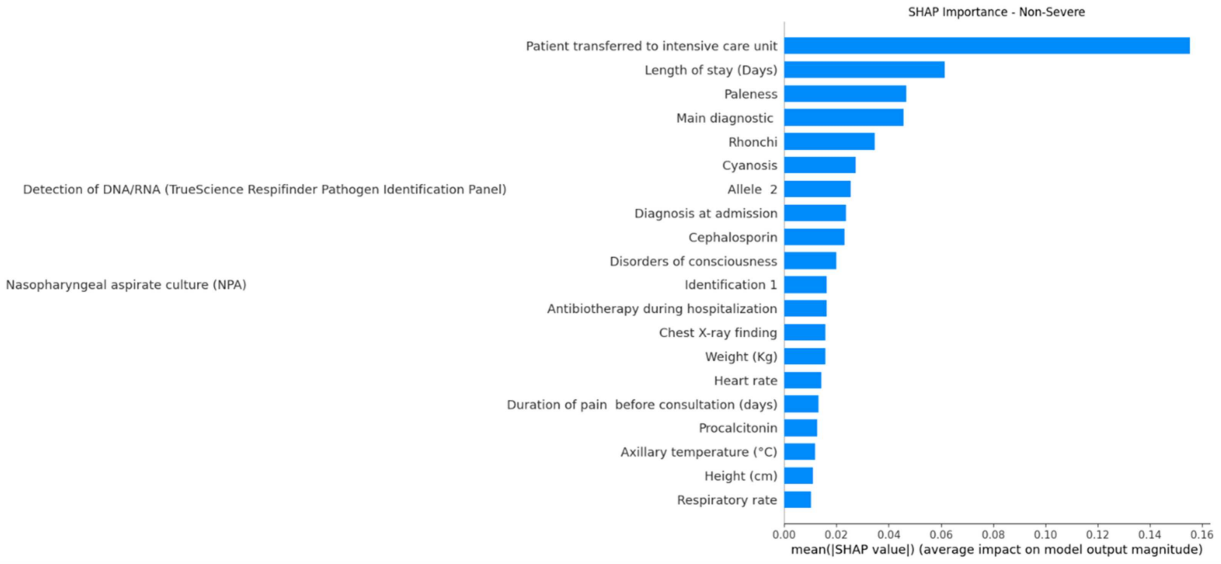
In Figure 8(a–c), these three pictures offer insights into the interpretability of a model that uses the SHAP approaches to

Figure 7
Beeswarm SHAP plot—Random Forest classifier (Class 0), (b) SHAP summary bar plot—Random Forest classifier (Class 0), (c) LIME plot—Random Forest classifier (Class 0)

(a)



(b)



(c)

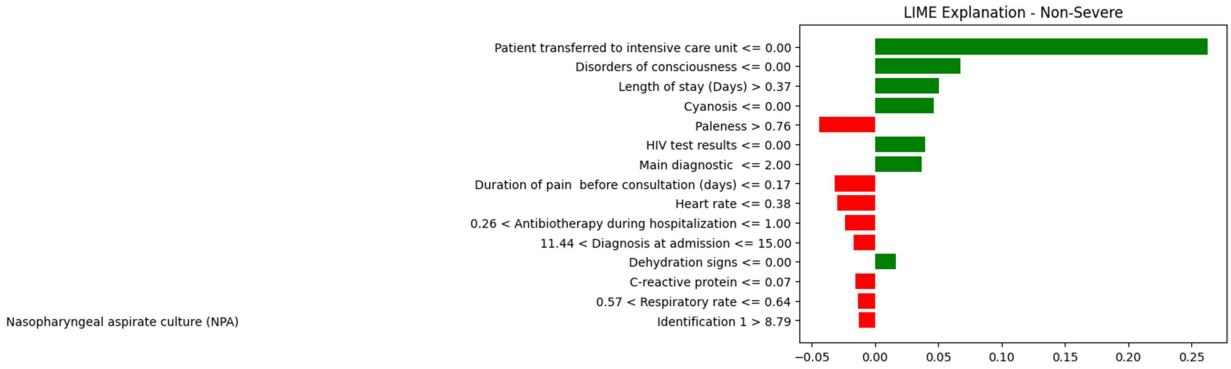
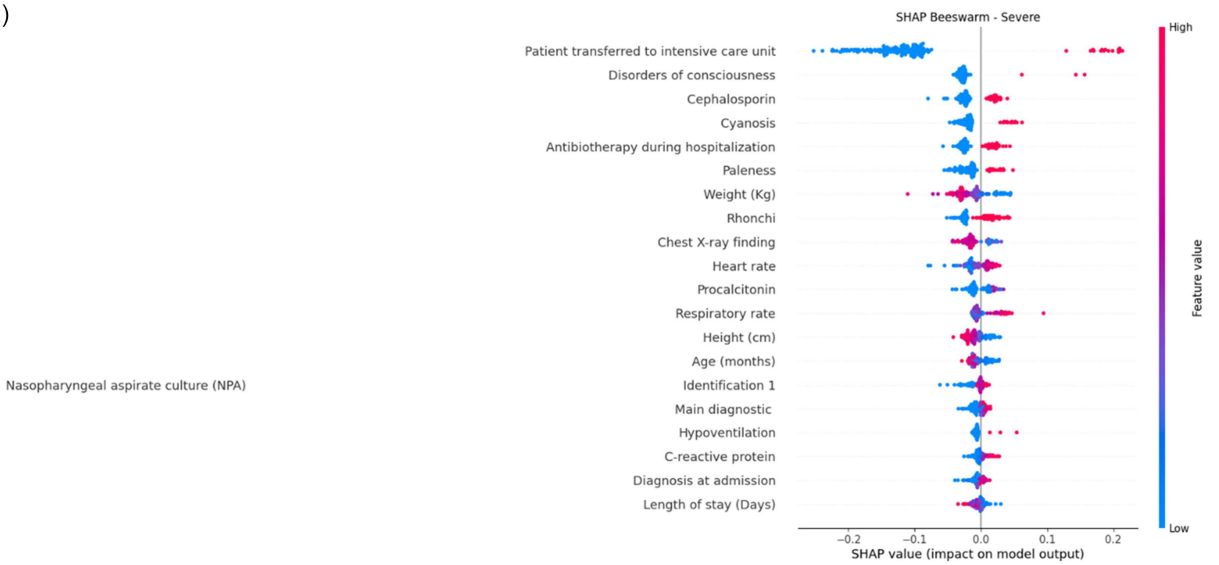


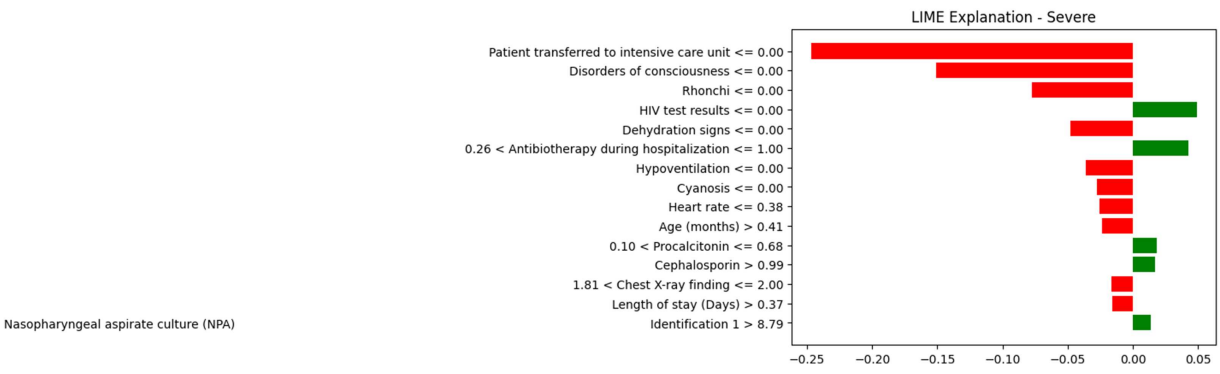
Figure 8
Beeswarm SHAP plot—Random Forest classifier (Class 1), (b) SHAP summary bar plot—Random Forest classifier (Class 1), (c) LIME plot—Random Forest classifier (Class 1)

(a)



(b)

(c)



predict severe patient outcomes (Class 1). A good explanation of how various characteristics affect an ML model's prediction for the "Severe" class in a medical dataset can be found in the two SHAP charts. Each dot in the first image, which is a Beeswarm SHAP plot, represents a distinct patient. The color denotes whether the feature value is low (blue) or high (red), and the horizontal position of the dot reflects the SHAP value (i.e., how much that feature pushes the prediction toward or away from the severe class). For instance, a shorter value of "patient transferred to intensive care unit" (blue dots on the left) is strongly linked to a reduced chance of being classified as "Severe."

The second image is a SHAP summary bar plot, which uses the mean of the absolute SHAP values to rank features according to their overall importance. This illustrates the average contribution of each feature, independent of direction, to the model. The most significant factor is the patient transferred to intensive care unit, which is followed by disorders of consciousness, cephalosporin, cyanosis, and antibiotherapy during hospitalization. Together, these plots offer a comprehensive picture of feature significance as well as a close-up look at how specific features affect model predictions for every scenario.

The 8(c) picture illustrates how particular feature values affect a single prediction and is a LIME explanation plot for the "Severe" class. Green bars show characteristics that drive the forecast away from the "Severe" classification, while red bars show characteristics that push it in that direction. A refusal of transfer to the critical care unit followed by disorders of consciousness, rhonchi, and dehydration signs are the main factors contributing to the "Severe" forecast in this instance. On the other hand, characteristics such as height, age, the existence of HIV test results, antibiotherapy during hospitalization, procalcitonin, and cephalosporin decrease the probability of being labeled as "Severe," as indicated by their green bars. A localized, instance-specific understanding of the model's classification decision-making process is provided by this LIME display.

The two visualizations of Figure 9(a) and (b) employ SHAP methodology, providing interpretability insights to predict the "At-Risk" class. Figure 9(a) Beeswarm plot, which shows how each feature affects the model's output for the "At-Risk" class, is the first image. Every dot stands for a patient, and its color (red for high, blue for low) and location on the x-axis (which represents the SHAP value; positive values improve risk prediction, negative values decrease it) indicate the feature value. High "length of stay (days)" and tend to push predictions toward the At-Risk class (left of zero), suggesting they are important factors in elevated risk.

The second image is a bar plot of the SHAP summary, which shows the relative contribution of each feature to the overall predictions of the model by ranking the features according to their average absolute SHAP value. With a mean SHAP value above 0.04, the leading contributors are "length of stay (days)," "rhonchi," and "main diagnostic." In contrast to the Beeswarm plot, this plot offers a more aggregated viewpoint, which facilitates comprehension of the global significance of variables throughout the dataset rather than just individual occurrences.

In Figure 9(c), the "At-Risk" class is shown in the third picture. By determining the influence of individual feature values on a particular prediction, LIME offers instance-specific interpretability. Green bars show features that pushed the forecast away from the "At-Risk" class, while red bars show features that had a negative influence on the "At-Risk" prediction. Disorders of consciousness, rhonchi, and heart rate worked against the risk estimate for this patient, but positive HIV results and a brief hospital stay significantly supported it.

The evaluation of various ML methods for pediatric respiratory disorders is given in Table 7. Our study surpasses the findings of prior studies using the same dataset of juvenile respiratory disorders by achieving the highest accuracy (99.29%) with advanced preprocessing and Random Forest model optimization. By tackling significant issues including class imbalances and inconsistent data that were overlooked in earlier studies, this method significantly increased the models' performance. A resilient model was created utilizing methods like data preprocessing, encoding, outlier removal, oversampling, and feature selection to establish the benchmark for predicting the survival rate of pediatric respiratory illnesses.

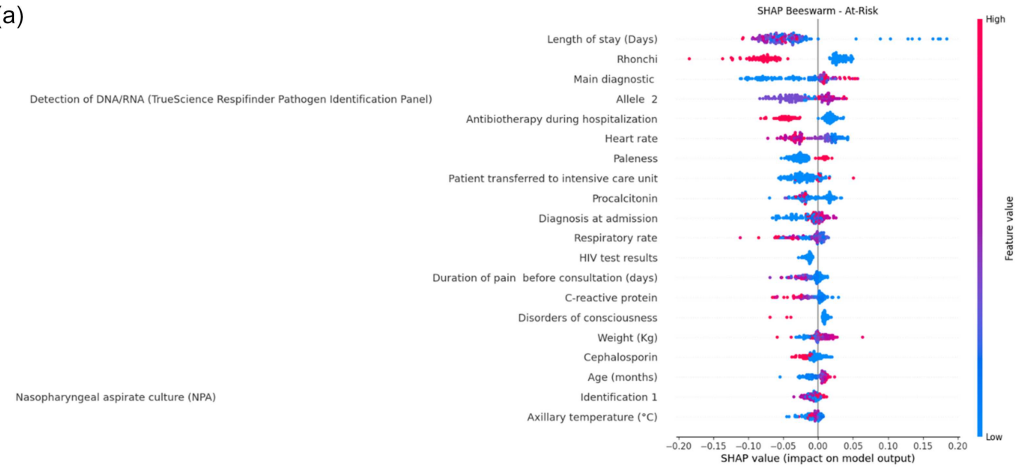
As shown in Table 8, the proposed model is significantly better than existing methods in terms of prediction accuracy and clinical applicability. The proposed Random Forest model, enhanced with XAI integration, achieved an outstanding accuracy of 99.29%, which is higher than any previous research, while earlier studies reported accuracies between 70% and 97%. The model employed 31 carefully chosen clinical variables, thereby avoiding the redundancy and noise often associated with high-dimensional datasets, such as the 73 features used by Mario et al., and it provided more comprehensive data than studies using fewer features, for example, 9, 15, or 16. This enhancement can be attributed to an optimal balance in feature selection. Although Kumar et al. also used Random Forest with XAI, their accuracy of 96% was considerably lower than the proposed model, underscoring the benefit of the refined feature set and improved model architecture. Besides, the diminished sample size in our study (801 instances) relative to Kumar et al. (1188 instances) results from more rigorous outlier elimination and data cleansing protocols to guarantee superior data quality. Additionally, the proposed method demonstrates that feature engineering and dataset quality can be more effective than sheer size, unlike Haile et al., who achieved only 72% accuracy despite using a very large dataset of over 327,000 cases. The recommended method provides superior accuracy and ensures interpretability by combining ensemble learning with understandable insights. This allows physicians to better understand the importance of features and have greater confidence in the system's predictions. Due to its performance, balanced use of features, and transparency, the proposed model is more robust and clinically reliable than previous studies.

6. Conclusion

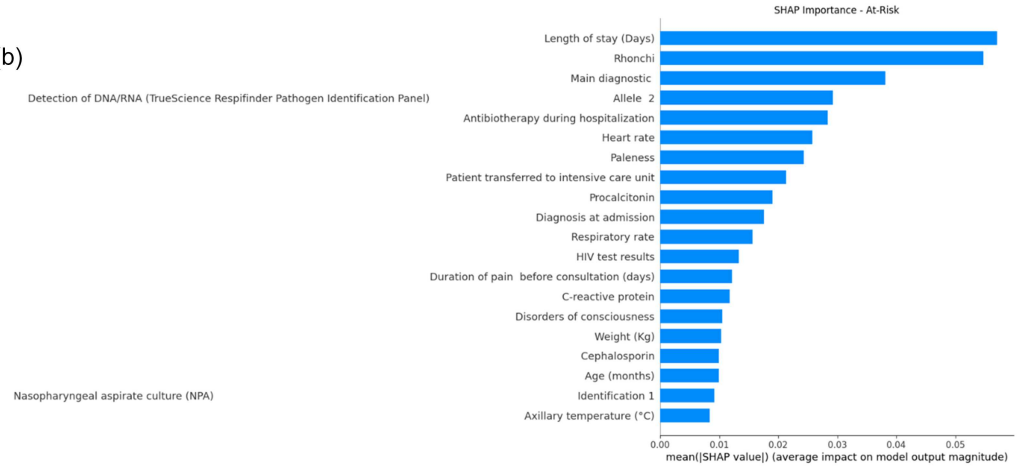
This research enhances pediatric critical care by demonstrating that Random Forest can accurately predict ICU mortality in respiratory diseases with 99.29% accuracy. Data reliability was ensured through thorough preprocessing, including SMOTE-ENN, hybrid feature selection, and outlier removal. It enhanced model interpretability by using SHAP and LIME, which fostered clinician confidence and increased the model's dependability and trustworthiness. By exceeding earlier benchmarks, the study establishes a new standard for predictive modeling. Future research will incorporate deeper learning, neural networks, additional ML classifiers, and the integration of multimodal data, such as medical imaging, along with diverse demographics, to improve precision, recall, F1-score, and specificity. These advancements will boost ML-driven diagnoses, improve outcomes for pediatric patients, and address real-world healthcare requirements. Future research can improve pediatric mortality prediction even further by looking at deep learning architectures such as recurrent and convolutional

Figure 9
Beeswarm SHAP plot—Random Forest classifier (Class 2), (b) SHAP summary bar plot—Random Forest classifier (Class 2), (c) LIME plot—Random Forest classifier (Class 2)

(a)



(b)



(c)

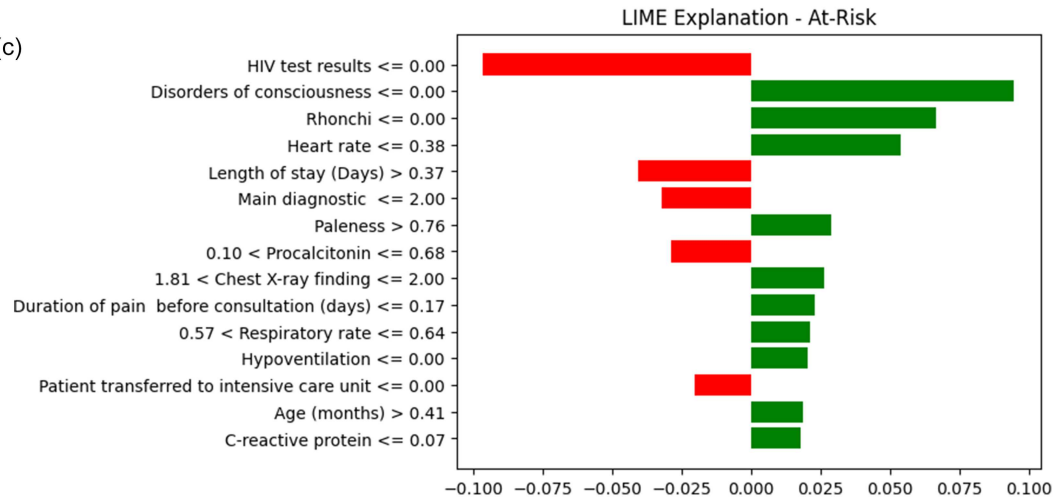


Table 8
Assessing various machine learning approaches to predict the survivability rate of pediatric respiratory

Authors	Dataset used	No. of instances	Classification	Accuracy (%)
Prithula et al. [4]	Clinical data (801 patients)	16	CatBoost (subdivision technique)	89.32
Kumar et al. [1]	Clinical data (1188 patients)	15	Random Forest (with XAI)	96.00
Chak et al. [38]	Clinical data (700 patients)	9	Extreme Gradient Boosting	91.9
Mario et al. [39]	Clinical data (365 patients)	73	Random Forest	97
Jorge et al. [40]	Clinical data	7	K-Nearest Neighbor	70
Proposed model	Clinical data (801 patients)	32	Random Forest (with XAI)	99.29

neural networks, which may be better able to understand nonlinear relationships and temporal patterns in clinical data. Integrating multimodal data sources, such as genomic data, electronic health records, and medical imaging, can improve prediction accuracy and provide a more complete picture of patient health.

7. Future Work

Future studies can improve pediatric mortality prediction even more by investigating deep learning architectures like recurrent and convolutional neural networks, which may be better able to grasp nonlinear relationships and temporal patterns in clinical data. Prediction accuracy may be increased, and a more comprehensive picture of patient health may be obtained by the integration of multimodal data sources, such as genomic data, electronic health records, and medical imaging. Larger, multicenter datasets would also improve model generalizability and lessen the bias that comes with single-institution research. Besides, future research should evaluate the model on external multicenter datasets to ensure its applicability across various hospital settings.

Ethical Statement

The dataset used in this study was publicly available and anonymized children's respiratory disease data from the Mendeley Data repository that were originally collected at the Children's Hospital of Rabat in Morocco. No personally identifiable patient information was used by the authors, and the data were de-identified clinical and demographic data. Ethical approval and informed consent were not required as this is a secondary data analysis of anonymized and publicly available data that did not involve direct contact with human participants. Ethical principles and data protection laws were followed in the study.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in the Mendeley Data repository at https://data.mendeley.com/datasets/2d9dvnycjw/1.

Author Contribution Statement

Khandaker Mohammad Mohi Uddin: Conceptualization, Methodology, Writing – review & editing, Supervision. **Anjuman Ansary:** Software, Validation, Formal analysis, Investigation, Writing – original draft. **Nazim Uddin:** Visualization, Project administration. **Md. Tofael Ahmed Bhuiyan:** Resources, Data curation.

References

[1] Kumar, R., Srirama, V., Chadaga, K., Muralikrishna, H., Sampathila, N., Prabhu, S., & Chadaga, R. (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

[2] Gan, H., Hou, X., Zhu, Z., Xue, M., Zhang, T., Huang, Z., . . . , & Sun, B. (2022). Smoking: A leading factor for the death of chronic respiratory diseases derived from Global Burden of Disease Study 2019. *BMC Pulmonary Medicine*, *22*(1), 1–11. https://doi.org/10.1186/s12890-022-01944-w

[3] Momtazmanesh, S., Moghaddam, S. S., Ghamari, S. H., Rad, E. M., Rezaei, N., Shobeiri, P., . . . , & Ibitoye, S. E. (2023). Global burden of chronic respiratory diseases and risk factors, 1990–2019: An update from the Global Burden of Disease Study 2019. *EClinicalMedicine*, *59*, 101936. https://doi.org/10.1016/j.eclinm.2023.101936

[4] Prithula, J., Chowdhury, M. E., Khan, M. S., Al-Ansari, K., Zughaier, S. M., Islam, K. R., & Alqahtani, A. (2024). Improved pediatric ICU mortality prediction for respiratory diseases: Machine learning and data subdivision insights. *Respiratory Research*, *25*(1), 1–16. https://doi.org/10.1186/s12931-024-02753-x

[5] Eisenberg, M. A., & Balamuth, F. (2022). Pediatric sepsis screening in US hospitals. *Pediatric Research*, *91*(2), 351–358. https://doi.org/10.1038/s41390-021-01708-y

[6] Quadir, A., Festa, M., Gilchrist, M., Thompson, K., Pride, N., & Basu, S. (2024). Long-term follow-up in pediatric intensive care–A narrative review. *Frontiers in Pediatrics*, *12*, 1430581. https://doi.org/10.3389/fped.2024.1430581

[7] Balamuth, F., Scott, H. F., Weiss, S. L., Webb, M., Chamberlain, J. M., Bajaj, L., . . . , & Alpern, E. R. (2022). Validation of the pediatric sequential organ failure assessment score and evaluation of third international consensus definitions for sepsis and septic shock definitions in the pediatric emergency department. *JAMA Pediatrics*, *176*(7), 672–678. https://doi.org/10.1001/jamapediatrics.2022.1301

[8] Sanford Kobayashi, E., Waldman, B., Engorn, B. M., Perofsky, K., Allred, E., Briggs, B., . . . , & Coufal, N. G. (2022). Cost efficacy of rapid whole genome sequencing in the pediatric intensive care unit. *Frontiers in Pediatrics*, *9*, 809536. https://doi.org/10.3389/fped.2021.809536

[9] Yu, G., Yu, Z., Shi, Y., Wang, Y., Liu, X., Li, Z., . . . , & Shu, Q. (2021). Identification of pediatric respiratory diseases using a fine-grained diagnosis system. *Journal of Biomedical Informatics*, *117*, 103754. https://doi.org/10.1016/j.jbi.2021.103754

[10] Papakyritsi, D., Iosifidis, E., Kalamitsou, S., Chorafa, E., Volakli, E., Peña-López, Y., . . . , & Roilides, E. (2023). Epidemiology and outcomes of ventilator-associated events in critically ill children: Evaluation of three different definitions. *Infection Control & Hospital Epidemiology*, *44*(2), 216–221. https://doi.org/10.1017/ice.2022.97

[11] Obajuluwa, A. M., Balogun, J. A., & Balogun, O. (2026). Fundamentals of radiology for physical therapists in low- and middle-income countries. In J. A. Balogun (Ed.), *Contemporary and global perspectives in physical therapy* (pp. 1813–1843). Springer Cham. https://doi.org/10.1007/978-3-031-86558-9_34

[12] Hucko, A., Weigel, R., Nizamov, F., & Black, M. (2025). Child and adolescent mortality in the WHO European Region: Concerning trends requiring urgent action. *Public Health in Practice*, *10*, 1–7. https://doi.org/10.1016/j.puhip.2025.100668

[13] Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, *263*, 110273. https://doi.org/10.1016/j.knosys.2023.110273

[14] Lötsch, J., Kringel, D., & Ultsch, A. (2021). Explainable artificial intelligence (XAI) in biomedicine: Making AI decisions trustworthy for physicians and patients. *BioMedInformatics*, *2*(1), 1–17. https://doi.org/10.3390/biomedinformatics2010001

[15] Gonem, S., Janssens, W., Das, N., & Topalovic, M. (2020). Applications of artificial intelligence and machine learning in respiratory medicine. *Thorax*, *75*(8), 695–701. https://doi.org/10.1136/thoraxjnl-2020-214556

[16] Patel, D., Hall, G. L., Broadhurst, D., Smith, A., Schultz, A., & Foong, R. E. (2022). Does machine learning have a role in the prediction of asthma in children? *Paediatric Respiratory Reviews*, *41*, 51–60. https://doi.org/10.1016/j.prrv.2021.06.002

[17] Dumbuya, J. S., Zeng, C., Deng, L., Li, Y., Chen, X., Ahmad, B., & Lu, J. (2025). The impact of rare diseases on the quality of life in paediatric patients: Current status. *Frontiers in Public Health*, *13*, 1531583. https://doi.org/10.3389/fpubh.2025.1531583

[18] Ganatra, H. A. (2025). Machine learning in pediatric healthcare: Current trends, challenges, and future directions. *Journal of Clinical Medicine*, *14*(3), 807. https://doi.org/10.3390/jcm14030807

[19] Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, *30*, 100924. https://doi.org/10.1016/j.imu.2022.100924

[20] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

[21] Lee, S. W., Loh, S. W., Ong, C., & Lee, J. H. (2019). Pertinent clinical outcomes in pediatric survivors of pediatric acute respiratory distress syndrome (PARDS): A narrative review. *Annals of Translational Medicine*, *7*(19), 513. https://doi.org/10.21037/atm.2019.09.32

[22] Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, *18*(1), 271. https://doi.org/10.3390/ijerph18010271

[23] Spathis, D., & Vlamos, P. (2019). Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health Informatics Journal*, *25*(3), 811–827. https://doi.org/10.1177/1460458217723169

[24] Aykanat, M., Kılıç, Ö., Kurt, B., & Saryal, S. B. (2020). Lung disease classification using machine learning algorithms. *International Journal of Applied Mathematics Electronics and Computers*, *8*(4), 125–132. https://doi.org/10.18100/ijamec.799363

[25] Chang, T. H., Liu, Y. C., Lin, S. R., Chiu, P. H., Chou, C. C., Chang, L. Y., & Lai, F. P. (2023). Clinical characteristics of hospitalized children with community-acquired pneumonia and respiratory infections: Using machine learning approaches to support pathogen prediction at admission. *Journal of Microbiology, Immunology and Infection*, *56*(4), 772–781. https://doi.org/10.1016/j.jmii.2023.04.011

[26] Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An evaluation of discriminative power in credit risk. *Frontiers in Artificial Intelligence*, *4*, 752558. https://doi.org/10.3389/frai.2021.752558

[27] Shah, N., Arshad, A., Mazer, M. B., Carroll, C. L., Shein, S. L., & Remy, K. E. (2023). The use of machine learning and artificial intelligence within pediatric critical care. *Pediatric Research*, *93*(2), 405–412. https://doi.org/10.1038/s41390-022-02380-6

[28] Ma, Y., Liu, W., Huang, B., Tian, F., & Chen, G. (2024). EFGAN: An automatic GAN-based methodology for mining eddy-front coupling with fused remote sensing data. *Information Fusion*, *101*, 101982. https://doi.org/10.1016/j.inffus.2023.101982

[29] Allgaier, J., & Pryss, R. (2024). Cross-validation visualized: A narrative guide to advanced methods. *Machine Learning and Knowledge Extraction*, *6*(2), 1378–1388. https://doi.org/10.3390/make6020065

[30] Bowen, D., & Ungar, L. (2020). Generalized SHAP: Generating multiple types of explanations in machine learning. *arXiv Preprint: 2006.07155*.

[31] Tligui, H., Hattoufi, K., Jroundi, I., Mahraoui, C., & Bassat, Q. (2024). Pediatric respiratory infections: Epidemiological and etiological data from a cohort of 801 Moroccan children. [Data set]. Mendeley Data. https://doi.org/10.1016/j.dib.2024.110457

[32] Perikova, E. I., Filippova, M. G., Makarova, D. N., & Gnedykh, D. S. (2024). The benefits of labeling in fast mapping and explicit encoding. *Neuroscience and Behavioral Physiology*, *54*(3), 424–433. https://doi.org/10.1007/s11055-024-01609-7

[33] Elkari, B., Chaibi, Y., & Kousksou, T. (2024). Random forest with feature selection and K-fold cross validation for predicting the electrical and thermal efficiencies of air based photovoltaic-thermal systems. *Energy Reports*, *12*, 988–999. https://doi.org/10.1016/j.egyr.2024.07.002

[34] Talan, T. (2026). Effectiveness of machine learning methods in predicting water potability: A comparative study. *Water Resources Management*, *40*(1), 25. https://doi.org/10.1007/s11269-025-04421-1

[35] Abdumalikov, S., Kim, J., & Yoon, Y. (2024). Performance analysis and improvement of machine learning with various feature selection methods for EEG-based emotion classification. *Applied Sciences*, *14*(22), 10511. https://doi.org/10.3390/app142210511

[36] Iranzad, R., & Liu, X. (2025). A review of random forest-based feature selection methods for data science education and applications. *International Journal of Data Science and Analytics*, *20*(2), 197–211. https://doi.org/10.1007/s41060-024-00509-w

[37] Cheng, H., Wen, Z., Zhao, Y., Wu, Z., Ren, X., & Yue, Z. (2024). Effect and optimization of geometric parameters and arrangement on film cooling performance of fan-shaped holes based on generalized regression neural network. *International Communications in Heat and Mass Transfer*, *158*, 107868. https://doi.org/10.1016/j.icheatmasstransfer.2024.107868

[38] Tso, C. F., Lam, C., Calvert, J., & Mao, Q. (2022). Machine learning early prediction of respiratory syncytial virus in pediatric hospitalized patients. *Frontiers in Pediatrics*, *10*, 886212. https://doi.org/10.3389/fped.2022.886212

[39] Lovrić, M., Banić, I., Lacić, E., Pavlović, K., Kern, R., & Turkalj, M. (2021). Predicting treatment outcomes using explainable machine learning in children with asthma. *Children*, *8*(5), 376. https://doi.org/10.3390/children8050376

[40] Amaral, J. L., Lopes, A. J., Jansen, J. M., Faria, A. C., & Melo, P. L. (2012). Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease. *Computer Methods and Programs in Biomedicine*, *105*(3), 183–193. https://doi.org/10.1016/j.cmpb.2011.09.009

How to Cite: Uddin, K. M. M., Ansary, A., Uddin, N., & Bhuiyan, MT. A. (2026). Explainable AI-Driven Clinical Decision Support for Mortality Risk Stratification in Pediatric Respiratory Disorders. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62028365