



ARTICLE

Enhancing Septic Shock Detection through Interpretable Machine Learning

Md Mahfuzur Rahman^{1,*}, Md Solaiman Chowdhury², Mohammad Shorfuzzaman³, Lutful Karim⁴,
Md Shafiullah⁵ and Farag Azzedin¹

¹Department of Information and Computer Science, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia

²Department of Electrical and Computer Engineering, North South University, Dhaka, 1229, Bangladesh

³Department of Computer Science, Taif University, Taif, 21974, Saudi Arabia

⁴College of Applied Arts and Technology, Seneca Polytechnic, Toronto, M2J 2X5, Canada

⁵Department of Control & Instrumentation Engineering, King Fahd University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia

*Corresponding Author: Md Mahfuzur Rahman. Email: mdmahfuzur.rahman@kfupm.edu.sa

Received: 15 June 2024 Accepted: 09 September 2024 Published: 31 October 2024

ABSTRACT

This article presents an innovative approach that leverages interpretable machine learning models and cloud computing to accelerate the detection of septic shock by analyzing electronic health data. Unlike traditional methods, which often lack transparency in decision-making, our approach focuses on early detection, offering a proactive strategy to mitigate the risks of sepsis. By integrating advanced machine learning algorithms with interpretability techniques, our method not only provides accurate predictions but also offers clear insights into the factors influencing the model's decisions. Moreover, we introduce a preference-based matching algorithm to evaluate disease severity, enabling timely interventions guided by the analysis outcomes. This innovative integration significantly enhances the effectiveness of our approach. We leverage a clinical health dataset comprising 1,552,210 Electronic Health Records (EHR) to train our interpretable machine learning models within a cloud computing framework. Through techniques like feature importance analysis and model-agnostic interpretability tools, we aim to clarify the crucial indicators contributing to septic shock prediction. This transparency not only assists healthcare professionals in comprehending the model's predictions but also facilitates the integration of our system into existing clinical workflows. We validate the effectiveness of our interpretable models using the same dataset, achieving an impressive accuracy rate exceeding 98% through the application of oversampling techniques. The findings of this study hold significant implications for the advancement of more effective and transparent diagnostic tools in the critical domain of sepsis management.

KEYWORDS

Sepsis prediction; machine learning; cloud computing

1 Introduction

Septic shock presents a complicated medical challenge triggered by infection within the human body, activating Systemic Inflammatory Reaction disorder (SIRs) [1]. This complex condition exerts a



considerable financial burden, accounting for approximately \$24 billion annually in the US healthcare system [2]. Studies indicate the life-threatening nature of sepsis, with mortality rates ranging from 25 to 40 percent [2,3]. Characterized by organ dysfunction resulting from a systemic inflammatory response to infection, sepsis manifests diversely among patients, posing challenges for swift identification [4]. Escalating at a rate of 11.9% per year, the associated costs highlight the urgency of ongoing research to establish timely surveillance methods, particularly through automated systems capable of detecting septic shock in hospitalized patients. A key obstacle in the early detection of sepsis lies in its subtle progression, with clinical symptoms proving challenging to discriminate in the initial stages. Early sepsis indicators lack specificity, complicating rapid identification. The integration of an automated clinical prediction and decision-making approach, leveraging Electronic Health Records (EHR), emerges as a promising strategy to enhance efficiency and broaden the spectrum of treatment for complex disorders [5]. This predictive methodology offers versatile applications, aiding healthcare professionals in measuring the severity of a patient's condition and monitoring their health status. The challenges intensify during medical emergencies involving a rush in hospitalizations, particularly in cases of critical heart problems. Such scenarios demand a precise risk assessment for each patient, necessitating various diagnostic techniques. The pressure on healthcare systems during these surges is profound, emphasizing the need for automated processes to mitigate human error and enhance overall efficiency. In this context, the prediction of septic shock stands to gain substantial benefits from the convergence of early identification using EHR and the application of a preference-matching algorithm for optimal treatment [6]. By integrating these components, healthcare professionals can streamline decision-making, ensuring timely interventions and personalized treatments based on individual patient profiles. The automated approach not only addresses the challenges posed by the delicate progression of sepsis but also facilitates a proactive and systematic response to the dynamic healthcare landscape. As research in this domain advances, the interaction between technological innovations and medical expertise holds the promise of significantly improving results for patients at risk of septic shock. Our study stands out from previous research efforts by taking a comprehensive approach to sepsis management. This involves not only early detection with interpretable results but also includes severity prediction and personalized treatment recommendations. In recent years, machine learning has significantly impacted healthcare, particularly in managing complex conditions like sepsis, where early detection is crucial for improving patient outcomes. Our study enhances septic shock detection by employing Classification and Regression Trees (CART) and addressing class imbalance with oversampling techniques like random oversampling, SMOTE, and ADASYN. By focusing on interpretable models, we aim to build trust and ensure seamless clinical integration, contributing to a more dynamic, transparent, and personalized approach to sepsis management. This study centers on the utilization of interpretable machine learning models within a cloud computing framework for the early detection of septic shock, aiming to address the critical need for transparency in decision-making processes. Interpretable machine learning, characterized by models whose predictions are easily understood and explained by humans, plays a crucial role in building trust and encouraging the use of advanced algorithms in clinical settings. The application of interpretable ML in septic shock detection represents a key shift in the diagnostic landscape, providing clinicians with insights into the factors influencing predictions and enabling more informed decision-making. The focus on interpretable machine learning models ensures that predictions are transparent and understandable to clinicians, fostering trust and facilitating adoption in clinical settings [7,8]. We leverage an extensive dataset comprising 1,552,210 EHRs, offering an unprecedented scale for our analysis. While interpretable machine learning and oversampling techniques are well-established methodologies, our study uniquely validates their efficacy, highlighting the exceptional performance of oversampling in sepsis prediction, surpassing a remarkable 98%. An innovative element of our research is the introduction

of a preference-based matching algorithm, which quantifies disease severity and makes treatment decisions based on individual patient preferences. The proposed approach integrates machine learning models, a vast dataset from Intensive Care Unit (ICU) patients, and a rigorous evaluation process, introducing the pioneering concept of preference-based matching in sepsis management. By incorporating patient-specific preferences and deploying diverse machine-learning algorithms such as Logistic Regression (LR), Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), Naive Bayes (NB), and Gaussian Mixture (GM), we conduct a thorough analysis of the dataset. To ensure a balanced dataset, we implemented a variety of undersampling and oversampling techniques in the training of our models. This approach yields superior predictions for septic shock, showcasing significant quantitative enhancements over prevailing state-of-the-art models. Notably, the CART model consistently outperforms others, achieving commendable accuracy (0.98), precision (0.98), and recall (0.98) with data oversampling. These outcomes represent a substantial leap forward in septic shock prediction accuracy, a crucial factor for enabling early intervention and ultimately improving patient outcomes. Furthermore, we applied the shapley analysis technique [9] in our investigation. Employing shapley analysis gave us a deeper understanding of how individual features contribute to and influence our model's predictions or classifications, enabling more informed decision-making. This helped us grasp the relative significance of each feature in shaping the outcomes of our model.

In summary, we have made the following substantial contributions to this paper:

- An innovative approach leveraging interpretable machine learning models and cloud computing for expedited septic shock detection through electronic health data analysis.
- Integration of state-of-the-art machine learning algorithms with interpretability techniques, providing accurate predictions with clear insights into the influencing features.
- Introduction of a preference-based matching algorithm to assess disease severity, enabling timely interventions guided by analysis outcomes.
- Application of feature importance analysis and model-agnostic interpretability tools to elucidate critical indicators contributing to septic shock prediction.
- Validation of the interpretable models' efficacy through a large EHR dataset, achieving an impressive accuracy rate exceeding 98% with oversampling techniques.

The rest of the paper is organized as follows. [Section 2](#) provides an overview of the existing techniques through a literature review. [Section 3](#) delves into the proposed methodology with system architecture. [Section 4](#) presents the experiment results and facilitates discussion, while the paper concludes with some future work in [Section 5](#).

2 Related Works

Ghosh et al. [1] used machine learning algorithms to predict septic shock early and investigated its performance. The method combines highly informative sequential patterns with a coupled hidden Markov model and several physiological variables to capture the interactions between those patterns. Moreover, their experiment used heart rate, mean arterial pressure, and respiratory rate. The authors asserted that their model can predict the occurrence of septic shock with high accuracy. They reported that SVN-MAPP had one of the greatest model accuracies, which was 77.2% for round 1, 82.1% for round 2, and 78.3% for round 3. In another study [10], the authors proposed a solution that uses a machine learning algorithm to determine sepsis mostly 48 h earlier using six vital signs. The gradient boosting approach was utilized to create the classifier. The dataset they collected from two

different sources contains patients who were admitted to the hospital without having sepsis but having at least one of the six signs. They used an iterative approach to calculate the total score which was taken from the ensemble. Their focus was on the Area Under the Receiver Operating Characteristic (AUROC) curve to measure the accuracy, and the result was 0.89. This score played a vital role in determining the presence of sepsis on each patient. They claimed their experiment was successfully able to predict sepsis more accurately than commonly used tools. Kim et al. [11] demonstrated a prediction-based method for sepsis detection for those who are more than 20 years old and visit the emergency department. Some parameters were used to determine the outcome of the experiment and worked as predictors. The parameters they used in their experiment were vital signs, level of consciousness, chief complaints, and initial blood test results. According to the study, the total number of patients was 49,560; among them, 9.7% had septic shock within 24 h. Their experiment results showed that all the Machine Learning classifiers remarkably outperformed the qSOFA score [12]. Greco et al. [13] also explored machine learning models, including random forest, balanced, and unbalanced logistic regression, with traditional clinical scores such as qSOFA, SOFA, and APACHE II. Compared to traditional scores, ML models, especially random forest, demonstrate better accuracy in the early identification of high-risk septic patients, potentially improving outcomes.

Le et al. [14] suggested a new approach for predicting sepsis in the early stages. The UCSF Medical Center provided the dataset used in the investigation, and the data were collected between June 2011 and March 2016. The dataset contains EHRs of patients between the age of 2 to 17. According to their findings, 9486 patients were identified and 101 (1.06%) were judged to have serious sepsis. The authors assert that a machine learning algorithm (MLA) can dynamically monitor an electronic health record to detect critical sepsis and predict the likelihood of septic shock in pediatric inpatients. Additionally, an earlier diagnosis of sepsis may aid in initiating therapy before it progresses to any serious situation. In another study [15], Chiew et al. worked on a prediction-based system using a machine learning model to find the risk level of suspected sepsis victims in the emergency department. According to the authors, if they can identify the risk level of the sufferer in the emergency department before any severe situation, the emergency department may get sufficient time to make appropriate decisions and treatment; thereby, the result will be quite flawless. Patients' health records used in this experiment were taken from the Singapore General Hospital Emergency department from September 2014 to April 2016. The result from this study claimed that the best model was gradient boosting with a 0.50 F1 score. Fleuren et al. [16] proposed an advanced system with a machine-learning model. In this system, a promising real-time model was used to detect early clinical recognition of sepsis. They extracted 130 models from a total of 28 papers. According to their study, the papers they considered were developed in the intensive care unit, followed by hospital wards, the emergency department, and all of these settings. For the prediction of sepsis, their AUROC diagnostic accuracy scores ranged from 0.68 to 0.99 in the hospital and from 0.87 to 0.97 in the emergency department. Islam et al. [17] found that an early and precise sepsis forecast can assist doctors with legitimate medicines and help minimize the vulnerability. In this study, they conducted an electronic search for the studies that were associated with sepsis prediction using machine learning algorithms between 1st January 2000 to 1st March 2018. They found seven studies out of 135 that were appropriate to their ideas. They firmly stated that the machine learning model could perform much better than the existing general predicting system.

In [18], the authors used data mining techniques to predict septic shock early. In this work, Recent Temporal Patterns are utilized in the co-occurrence with the Support Vector Machine (SVM) classifier to construct a robust system that provides an interpretable demonstration for early sepsis prediction. This approach is linked to two separate prediction tasks. One is an early prediction at the visit level, while the other is an early prediction at the event level. The study found that the Recent

Temporal Pattern (RTP) based display outperforms all previous models for both determination tasks. Nemati et al. [19] proposed a prediction-based sepsis detection in the ICU using the machine learning model. Their results showed that the AUROC curve for both training and testing scored 0.85 for predicting sepsis 4 h in advance and the experiment was driven among 27,527 patients. The authors believe that sepsis is a leading cause of mortality and a factor in the high costs associated with treating patients with life-threatening situations. Early mediation with anti-microbials can be the chance of survival for septic patients. In any case, no approved framework is clinically available for real-time sepsis forecasting. A method for early septic shock detection utilizing an LSTM (Long Short-Term Memory) network was demonstrated by Fagerstrom et al. [20]. The primary focus of their experiment was to predict if a patient is going to develop septic shock when the patient is in the hospital, and this prediction result comes from the LSTM model. The data they used in their system has 59,000 admission records from the critical care units of a hospital between 2001 and 2012. The dataset's parameters include critical signs and symptoms, medical procedures, lab test reports, diagnoses, patient demographics, and mortality rate. The experiment employed Keras machine learning with a Google TensorFlow backend to create their "LiSep LSTM" system. They applied six-fold cross-validation to obtain accurate performance measurements for each network configuration. The authors concluded that the LiSep LSTM model has an AUROC of 0.83 and an Hours Before Onset (HBO) median of 48. Shimabukuro et al. [21] came up with a solution to reduce the average stay and mortality rate using a severe sepsis prediction system with the help of machine learning. They chose the intensive care units of the University of California, San Francisco Medical Center for their investigation. To estimate the average duration of stay, a randomized, controlled clinical trial was conducted under their supervision. Also, they observed the mortality rate in the hospital from December 2016 to February 2017. Results from 75 participants showed that 6 hospital mortality occurred among the 67 patients in the experimental group and 16 in the control group. They claimed that the machine learning-based predictor showed statistically significant results. The average length of hospital stay dropped from 13.0 to 10.3 days, a decrease of 20.6%. They also noticed a considerable change in the mortality rate. Surprisingly, the hospital mortality rate decreased from 21.3% to 8.96%. During this trial, no adverse incidents were discovered.

Bataille et al. [22] conducted a research where 100 critically ill patients with severe sepsis or septic shock, machine learning techniques were applied to transthoracic echocardiography (TTE) data to predict fluid responsiveness. The study found that several machine learning models, including partial least-squares regression (PLS) and neural network (NNET), were able to predict fluid responsiveness with accuracy comparable to the traditional method of assessing it through passive leg raising (PLR). Key echo-cardiographic parameters, such as inferior vena cava collapsibility, velocity-time integral, S-wave, E/Ea ratio, and E-wave, were identified as important factors for predicting fluid responsiveness. Choi et al. [23] addressed the limitation by evaluating the utility of continuous vital sign monitoring through wireless wearable devices coupled with machine learning analysis. While prior studies have demonstrated the importance of vital sign monitoring in sepsis detection, this work showcases substantial improvements in predictive performance, with AUROC values of up to 0.861, and notably, earlier detection of clinical deterioration by as much as 9 h compared to traditional manual measurements. This innovative approach not only advances the field of sepsis prediction but also underscores the potential of wearable technology to enhance early intervention strategies in critical care settings. Islam et al. [24] examined the critical role of ML and deep learning (DL) techniques in the early detection and prediction of sepsis using EHRs. Through a rigorous selection process, they identified and analyzed 42 relevant studies from a pool of 1942 articles. These studies predominantly employed retrospective approaches and spanned diverse geographic regions, with a

primary focus on the United States. Despite variations in datasets, sepsis definitions, and prevalence rates, ML/DL methods demonstrated promise in early sepsis prediction, leveraging longitudinal EHR data. This review underscores the significance of ML/DL techniques for enhancing sepsis detection and prediction, highlighting their potential to improve patient outcomes through the utilization of EHR data. [Table 1](#) summarizes the strengths and limitations of the discussed related works.

Table 1: Summary with the related works with their strength and limitations

Related research	Pros	Cons
Ghosh et al. [1]	<ul style="list-style-type: none"> -Combines sequential patterns and physiological variables for accurate prediction. -Achieved high model accuracies. 	<ul style="list-style-type: none"> -Specific feature requirements may limit generalizability. -Performance may vary depending on the dataset and patterns considered.
Barton et al. [10]	<ul style="list-style-type: none"> -Predicts sepsis 48 h earlier, allowing for early intervention. -Achieved a high AUROC curve accuracy. 	<ul style="list-style-type: none"> -Requires access to vital sign data, which may not be available in all healthcare settings. -Performance may depend on the choice of vital signs and dataset.
Kim et al. [11]	<ul style="list-style-type: none"> -Specifically tailored for the emergency department setting. -Machine learning classifiers outperform traditional qSOFA score. 	<ul style="list-style-type: none"> -Limited to predicting sepsis cases in the emergency department. -Requires specific features like vital signs and level of consciousness.
Le et al. [14]	<ul style="list-style-type: none"> -Focuses on early sepsis prediction in pediatric inpatients. -Addresses the need for pediatric-specific sepsis detection. 	<ul style="list-style-type: none"> -Limited to the pediatric age group. -Requires electronic health records, which may not be universally available.
Chiew et al. [15]	<ul style="list-style-type: none"> -Aims to determine the risk level of suspected sepsis victims. -Potential for timely decision-making in emergency departments. 	<ul style="list-style-type: none"> -Performance may vary depending on the choice of machine learning model. -F1 score of 0.50 indicates room for improvement.
Fleuren et al. [16]	<ul style="list-style-type: none"> -Explores real-time models for sepsis detection 	<ul style="list-style-type: none"> -Diagnostic accuracy may vary across settings.
Khoshnevisan et al. [18]	<ul style="list-style-type: none"> -Utilizes data mining techniques for early septic shock detection. -Achieves superior performance for early sepsis prediction. 	<ul style="list-style-type: none"> -May require expertise in data mining techniques. -Specific features and patterns may limit generalizability.
Nemati et al. [19]	<ul style="list-style-type: none"> -Focuses on sepsis detection in the ICU setting. -Achieves a high AUROC curve accuracy for early prediction. 	<ul style="list-style-type: none"> -Limited to the ICU context. -Performance evaluation may depend on the chosen prediction time frame.

(Continued)

Table 1 (continued)

Related research	Pros	Cons
Fagerstrom et al. [20]	<ul style="list-style-type: none"> -Utilizes LSTM network for early septic shock detection. -Achieves a respectable AUROC score. 	<ul style="list-style-type: none"> -Requires access to comprehensive patient admission records. -Performance may vary depending on LSTM model configuration.
Proposed study	<ul style="list-style-type: none"> -Utilizes machine learning and cloud computing for septic shock prediction. -Offers potential for real-time data analysis and personalized treatment recommendations. -Compares different machine learning models for prediction. -Effectiveness may depend on the quality and availability of data. 	<ul style="list-style-type: none"> -Deployment and integration of cloud-based systems may have associated challenges.

Zheng et al. [25] focused on addressing the challenge of predicting 28-day mortality in septic shock patients in the ICU. Introducing ShockSurv, a model developed with XGBoost, the study highlights its superior performance compared to existing methods. Emphasizing the importance of early prediction, the authors suggest that ShockSurv, utilizing clinically available data, has the potential to improve outcomes for septic shock patients in the ICU. Li et al. [26] addressed early sepsis detection in trauma patients using a machine-learning model. Focusing on ICU-admitted trauma cases, the study employs an XGBoost model, achieving an AUROC of 0.83 to 0.88. Outperforming logistic regression, the XGBoost model shows promise for personalized intervention and early treatment in critical trauma patients. van der Vegt et al. [27] systematically reviewed AI-based sepsis prediction algorithms, analyzing 30 articles in adult hospital settings. It identifies barriers, enablers, and decision points, mapping them to the SALIENT framework. Five studies show reduced mortality post-implementation. The SALIENT framework, adaptable to various AI tasks, validates an end-to-end implementation framework for healthcare AI. Alanazi et al. [28] investigated early sepsis detection using a machine-learning approach in adult ICU patients. Analyzing data from 1182 sepsis-diagnosed patients, the study employs regression and data mining models. The regression model emphasizes time, lactic acid, and temperature as significant factors. The study highlights the need for continuous improvement in sepsis prediction through refined data analysis for enhanced accuracy in clinical outcomes. Zhang et al. [29] created a real-time sepsis prediction model with high timeliness and interpretability. Using eight physiological indicators and the Local Interpretable Model-Agnostic Explanation (LIME) method based on Extremely Randomized Trees, the model achieves an AUROC above 0.76. Imbalance XGBoost shows high specificity (0.86). The model provides detailed predictions, aiding clinical workers, and offers a dynamic early warning for sepsis in critically ill patients, improving diagnostic efficiency and credibility. Li et al. [30] established and validated an ML model for predicting in-hospital mortality in sepsis-associated acute kidney injury patients, with the XGBoost model achieving an area under the curve (AUC) of 0.794 and demonstrating superior performance and interpretability through SHapley Additive exPlanations (SHAP) and LIME algorithms. Zhou et al. [31] developed a machine

learning model for predicting mortality in sepsis-associated acute kidney injury patients, featuring 15 critical variables. The categorical boosting algorithm, especially the CatBoost model, demonstrates superior predictive performance (ROC: 0.83), validated externally in two Chinese hospitals (ROC: 0.75). Bao et al. [32] provided machine learning models on extensive sepsis patient data, with the Light Gradient Boosting Machine (light GBM) demonstrating superior performance (AUC: 0.99 train set, 0.96 test set). The findings advocate for integrating the light GBM algorithm into clinical decision tools for enhanced prognosis and outcomes prevention. Pan et al. [33] developed machine learning models for early in-hospital mortality prediction in septic ICU patients using SOFA components, with Logistic Regression, Gaussian Naive Bayes, and Support Vector Machines models showing superior accuracy and positive net benefits in the 5%–50% probability threshold range. Li et al. [26] developed a machine learning model using Extreme Gradient Boosting to predict sepsis risk in ICU-admitted trauma patients. Extracting data from 4603 cases over 11 years, they engineered 485 features from 42 raw variables. The model exhibited strong performance (AUROC 0.83–0.88) across various prediction windows, outperforming a logistic regression counterpart. The study underscores the potential of timely sepsis identification in trauma patients, aiding personalized interventions and early treatment. Chen et al. [34] addressed postoperative sepsis as a major cause of mortality after liver transplantation (LT). Analyzing data from 786 LT recipients, the research developed and validated seven machine-learning models. The Random Forest Classifier (RF) model demonstrated superior performance, with an AUC of 0.731 and an accuracy of 71.6% in the internal validation set. Notably, factors such as red blood cell infusion, ascitic removal, blood loss, and anesthesia time were identified as crucial predictors of post-LT sepsis. The RF model's efficacy was further confirmed in an external validation set, emphasizing its potential to aid clinical decision-making for postoperative sepsis prevention in LT recipients.

This study introduces significant advancements in septic shock detection through the use of a preference-based matching algorithm and interpretable machine learning models. The preference-based matching algorithm tailors treatment recommendations based on the predicted severity level, ensuring personalized and timely care. Additionally, the emphasis on interpretable models like CART (Classification and Regression Trees) enhances transparency, allowing clinicians to understand the factors influencing predictions. These innovations combine personalized treatment with transparent analytics, improving predictive accuracy and fostering trust and integration into clinical practice. The contributions outlined in existing research highlight several advancements in septic shock detection using machine learning models. Despite these valuable contributions, several research gaps remain unaddressed in the current literature. Firstly, while the study emphasizes the integration of state-of-the-art machine learning algorithms, there is a need for deeper exploration into the specific interpretability methods employed and their comparative efficacy in revealing influential features. Secondly, while the research claims to elucidate critical indicators contributing to septic shock prediction through feature importance analysis, a deeper dive into the clinical relevance and real-world applicability of these identified indicators is warranted. The paper introduces a preference-based matching algorithm for assessing and reacting to disease severity and the assessment of its robustness across diverse healthcare settings or patient populations. Furthermore, investigation into the generalizability of the model across different healthcare institutions and datasets is essential. Achieving an impressive accuracy rate exceeding 98% with oversampling techniques on a large EHR dataset our study addresses these research gaps enhancing the understanding and applicability of the methodology in real clinical settings, ensuring its effectiveness and reliability in a diverse healthcare environment.

3 Methodology

The entire system was implemented using various machine-learning techniques. The approaches we employed in our experiment are listed below: a. Logistic Regression (LR), b. Linear Discriminant Analysis (LDA), c. Classification and Regression Trees (CART), d. Naive Bayes (NB), and e. Gaussian Mixture (GM). Logistic regression is a statistical analysis method that is used to predict an outcome based on earlier observations of a dataset [35]. In machine learning, LR is an important tool for analyzing datasets. With the help of this approach, an algorithm is being used in machine learning applications to classify data based on previous data. If more data appear in the dataset, the algorithm should better predict classifications within datasets. Using logistic regression, we can predict whether a particular scenario can happen or not. We used binary logistic regression in our research. Another analysis technique called LDA [36] has been used in this study. LDA is a way of making the supervised classification method for creating ML models. In a dataset, dimensionality reduction techniques are used to reduce the number of dimensions. Some applications like image recognition and predictive analysis use the LDA mode based on dimensionality reduction. For predictive modeling, Decision Trees are an important type of algorithm in machine learning. CART is a modern technique that is also known as the decision tree algorithm. The classical decision tree algorithm is one of the most popular algorithms for creating a model that predicts a certain value based on a target. Naive Bayes is a classification technique based on Bayes' Theorem for predictive modeling [37]. The model works with the independence of every input variable. This model is simple, easy to build, and very useful when dealing with large datasets. Even when working with highly sophisticated classification methods, the performance of NB is outstanding. Using different machine learning (ML) techniques serves several purposes aimed at improving predictive models' overall performance and robustness. Testing multiple models helps identify the one that best fits the data, achieving high accuracy, precision, recall, and F1 score, while also determining robustness against data variations. Techniques like Random Oversampling, SMOTE, and ADASYN address class imbalance by balancing class distribution. Various models offer insights into feature importance, with tools like SHAP enhancing interpretability. Each ML technique has strengths, such as decision trees handling non-linear data and logistic regression excelling in binary classification, and comparing them ensures manipulating these strengths. By rapidly processing and analyzing large real-world datasets, the system improves prediction timeliness and accuracy. It integrates patient preferences for personalized treatment recommendations and continuously updates prediction models with real-time data. The study highlights the superior performance of the CART model, showing high accuracy and precision, effectively addressing class imbalance. This synergy of advanced machine learning, cloud scalability, and personalized care significantly improves prediction accuracy and patient outcomes in septic shock management.

3.1 Dataset

In our system, we proposed a model to find out the best-suited results based on the patient's current medical status. For example, a patient may have various report results that might match the model from a previous record. This matching result will help the doctor identify the chances of having septic disease. Our feature model has around 17 features with 1,552,210 health records. Data we used in our experiment were taken from ICU patients in the United States of America and extracted from the 2019 PhysioNet Challenge [38]. The dataset carries 40,336 individual records. Each patient has one or more observations per hour after getting admitted into the ICU, and 40 different variables are present in each observation, including demographic data, vital signs, and laboratory tests. A total of 27,916 septic patients were found, accounting for 1.8% of the total 1,552,210 records. The number of remaining non-septic observations is 1,524,294. Applying different machine learning algorithms will

help us predict the chances of having septic disease. Before that, we used our data set to find out the accuracy and we found that the dataset satisfies more than 95% of accuracy in the machine learning model. In our dataset (refer to [Table 2](#)), we included various measurements or clinical parameters to build the machine learning model for sepsis prediction. These measurements capture different aspects of a patient's physiological constraints and laboratory values. Each measurement corresponds to a specific aspect of the patient's health, such as heart rate, temperature, blood pressure, respiratory rate, and various laboratory test results. The "ICULOS" represents the length of stay in the ICU in hours since admission. It provides information about the duration of the patient's stay in the intensive care unit, which can be relevant in sepsis prediction. The "SepsisLabel" column indicates whether the patient is classified as septic (1) or non-septic (0). This is the target variable or the outcome we aim to predict using the machine learning model.

Table 2: Measurements that we used in our dataset to build the machine learning model

Measurement	Description
HR	Heart rate (beats per minute)
Temp	Temperature (deg C)
MAP	Mean arterial pressure (mm Hg)
Resp	Respiration rate (breaths per minute)
BaseExcess	Excess bicarbonate (mmol/L)
FiO2	Fraction of inspired oxygen (%)
AST	Aspartate transaminase (IU/L)
BUN	Blood urea nitrogen (mg/dL)
Calcium	Calcium (mg/dL)
Creatinine	(mg/dL)
Bilirubin_direct	Direct bilirubin (mg/dL)
TroponinI	Troponin I (ng/mL)
HCT	Hematocrit (%)
WBC	Leukocyte count (count/L)
ICULOS	ICU length of stay (hours since ICU admission)
SepsisLabel	Septic (1)/Non-Septic (0)

3.2 System Architecture

Our developed system takes all the medical-related parameters to generate the prediction of septic diseases. This will help the doctor to identify the patients and help them to make the queue sorted according to the emergency level of the patients during any diagnosis. For that, we used a dataset with multiple organ attributes defining the result of each patient if he/she has a septic disease. Using this dataset, we can learn how to predict septic disease from our system. However, prior to that, it is necessary to examine our dataset in terms of accuracy, identifying any missing values, determining feature importance, and conducting an analysis. Our focus is to predict the chances of having the septic disease or not. So, before suggesting any treatment, this can help to predict sepsis in advance and identify sepsis onset more accurately than the traditional way. If left untreated, the likelihood of mortality is considerably elevated. Upon receiving the sepsis prediction outcome, we determine the

probability of encountering septic shock. Based on the results, we ascertain the patient's preferences. Our model operates in two distinct stages. Initially, we make predictions regarding the level of septic shock. Subsequently, our system assesses whether a patient requires immediate emergency care or moderate supervision. We employed various machine learning models to explore and visualize our dataset. The dataset we utilized holds the potential for training the model to accurately predict the sepsis level of a patient. These predictions aid in the early detection of patients, surpassing the efficiency of commonly employed methods. After detecting the patient, we pass the result in a selection model algorithm to determine which patients are in emergency need according to their health situations.

In Fig. 1, we have our main model flowchart and we have mentioned the pseudocode steps for preparing our models in Listing 1. The system architecture is designed to effectively manage patient health data for sepsis risk assessment and personalized decision-making. It begins with the gathering of diverse patient health data from various sources, followed by moderately complex data preprocessing tasks to ensure data quality. The selection of suitable machine learning models, requiring domain expertise, is a crucial step in handling healthcare data and predicting sepsis. Data splitting, model training, and evaluation are standard processes while addressing class imbalances. Model interpretability enables us to understand feature importance, which is crucial for prediction-based algorithms. It helps identify the key variables that heavily influence the outcome and aids in making accurate predictions. The risk-checking condition helps to identify whether a patient is at risk of having sepsis or not using the machine learning model. The preference is then considered for the patients in risk to select a proper treatment plan for them considering the severity of the sickness and their preferences to treat it. The system's innovation lies in incorporating preference parameters and enabling alignment with patient choices. The treatment selection algorithm, potentially complex, integrates preferences with clinical data, and patients are excluded from the emergency list when no sepsis risk is indicated. The final treatment selection process complexity varies with the sophistication of algorithms balancing patient preferences and clinical considerations, ultimately enhancing personalized healthcare decision-making.

Listing 1: Pseudocode for Model Training and Evaluation

- 1 Load and preprocess data
- 2 Split the dataset into training and testing sets
- 3 For technique in [Random, NearMiss, SMOTE, ADASYN]:
- 4 Apply technique to balance the dataset
- 5 Train and Evaluate model on resampled data
- 6 Measure Accuracy, Precision, Recall, F1 score, and AUC
- 7
- 8 Use SHAP to explain feature importance of model
- 9 Plot feature importance values

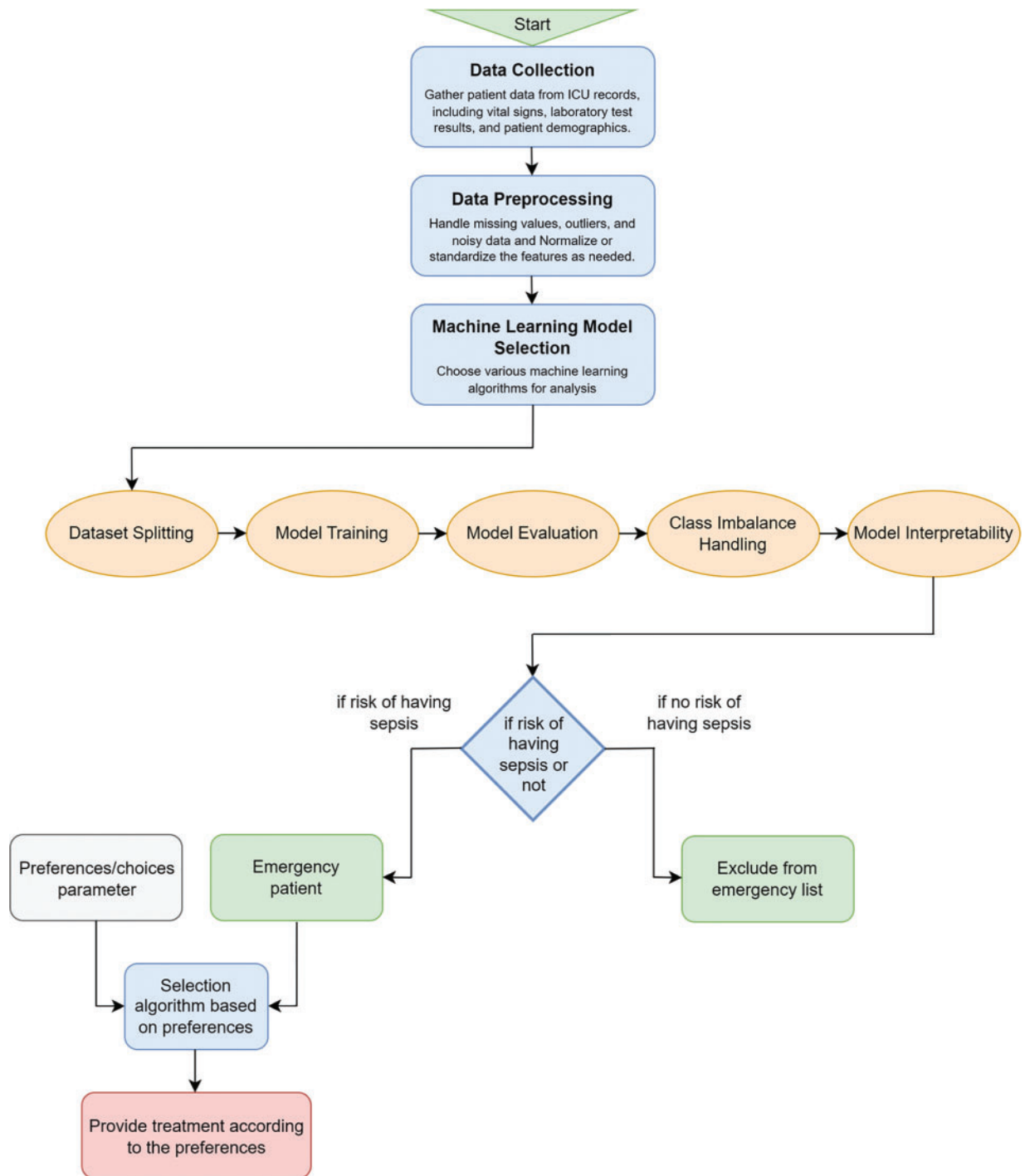


Figure 1: Proposed selection model according to the sepsis level prediction

In our strategy, we propose a cloud-based architecture that incorporates our machine learning model. We considered our previous work, described in [6,39], to develop a real-time healthcare application where the system makes the decision accurately for the patients and selects the best

treatment plan based on medical emergency and the preferences regarding treatment strategy. Fig. 2 describes how the best-suited hospital is selected based on the preferences of the patients where each hospital provides the services (with different qualities). Fig. 3 illustrates the interaction between the physician, patients, and our sepsis detection model, which is seamlessly integrated with a cloud system. The application considers several hospitals providing services to patients with different qualities (e.g., the success rate for treatment, current waiting time, user's review of the services, etc.). Each hospital has a certain capability to fulfill the patient's preference requirements. For example, a hospital can provide 24/7 emergency medical support, and on the other hand, it can have the capability to start major treatment immediately. Also, a hospital may not have sufficient ICUs. We consider all these as parameters of a hospital's capability level (using the mathematical model described in [6]). So, the hospital's medical service capability S_i can be described as, $S_i = [f_{i1} f_{i2} f_{i3} \dots f_{in}]$; where f_{ij} denotes the supply/quality level on a service/feature j of a hospital i . The application also considers two aspects related to the patient's preferences. One is the actual preference values (presented as V_i) and the other is the preferences' relative weights (presented as W_i). The total number of preference items is denoted as n , and the preference requirement of a patient P_i can be represented by (V_i, W_i) where $V_i = [v_{i1} v_{i2} v_{i3} \dots v_{in}]$ and $W_i = [w_{i1} w_{i2} w_{i3} \dots w_{in}]$. The V_{ij} specifies the passenger P_i 's preference value on the quality of the considered feature j , The W_{ij} denotes the priority or relative weight on feature j of the patient P_i (this weight value can be ignored if there's no preference requirement value available). Our cloud system embeds the machine learning model that first detects the sepsis and then suggests the best hospital for the patient treatment respecting the preferences and also the weights on the preferences. One of the primary reasons for adopting a cloud-based approach is scalability. Healthcare systems often experience fluctuations in demand, especially in emergency situations. Cloud computing allows us to instantly scale up or down, ensuring that our application can handle surges in patient data, ensuring uninterrupted service even during critical situations. This scalability directly translates into improved patient care by guaranteeing that our sepsis detection model and hospital selection algorithms remain responsive and accurate under any circumstances. Furthermore, the cloud's flexibility enables us to efficiently manage and store vast amounts of patient data securely. Data security and privacy are paramount in healthcare, and cloud providers invest heavily in cutting-edge security measures and compliance standards. This ensures that patient information remains confidential and protected against cyber threats. Our cloud-based solution also offers geographic distribution, making healthcare services more accessible to patients in remote areas. It enables patients to access our application from anywhere, at any time, breaking down geographical barriers and providing equitable healthcare access to all. Moreover, the cloud facilitates seamless collaboration between healthcare providers. Physicians can access patient data, share diagnostic insights, and collaborate on treatment plans in real time, leading to faster and more accurate decision-making. In terms of cost-effectiveness, cloud computing allows us to optimize resource utilization, reducing infrastructure costs and energy consumption. These cost savings can be redirected into improving healthcare services and expanding our capabilities.

4 Model Analysis

Several analyses have been made to identify the outcome of the process. We have computed scores from various models to compare the performance. We have considered the following list as our outcome measurement: a. Accuracy, b. Precision, c. Recall, d. F1 score. The accuracy of a machine learning classification algorithm is one way to estimate how often the algorithm classifies a data point correctly. By analyzing the accuracy results, we can determine the extent to which the predicted data points align with the actual data points, thereby providing an indication of the model's correctness in its predictions. When making a prediction from a model, we mostly rely on the accuracy result to

justify our outcome. Precision is one of the key factors that help predict a model’s accuracy. Precision, also known as a positive predictive value, refers to the proportion of relevant instances correctly identified among all the instances that were retrieved. For example, if a model avoids a lot of mistakes in predicting two classifications, then the model has high precision. Simply we can say that it is the ratio between true positives and all the positives. Recall is the ability of a model to find all the similar cases inside a dataset. Simply we can say that recall is how to correctly identify true positives in a model. From the mathematical understanding, recall is the true positives divided by the sum of true positives and false negatives. The F1 score is the average of the precision and recall. It combines both the precision and recall of a classifier into account and provides their harmonic mean. Sometimes it is hard to make a decision only from the accuracy score, and in that case, F1 score is more useful. Before conducting the process, it is necessary to check whether the dataset is balanced or imbalanced. Resampling methods can be a very good choice for an imbalanced classification task. Oversampling duplicates synthetic examples in minority class for an imbalanced classification. On the other hand, undersampling techniques delete instances in the majority class.

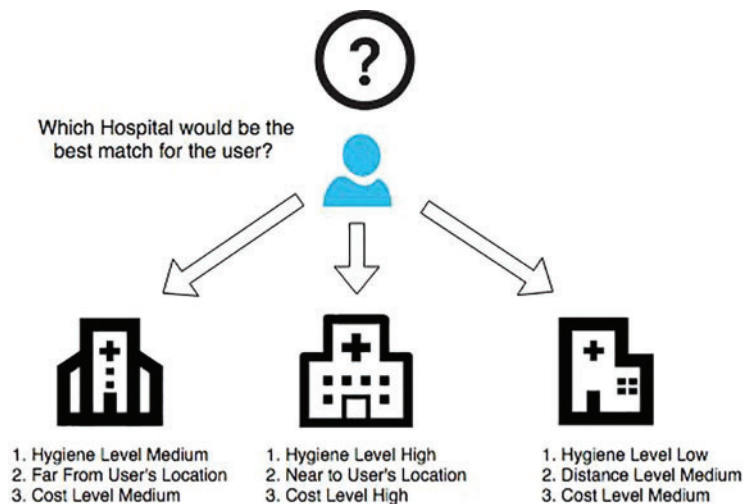


Figure 2: Patient’s preference-based hospital matching

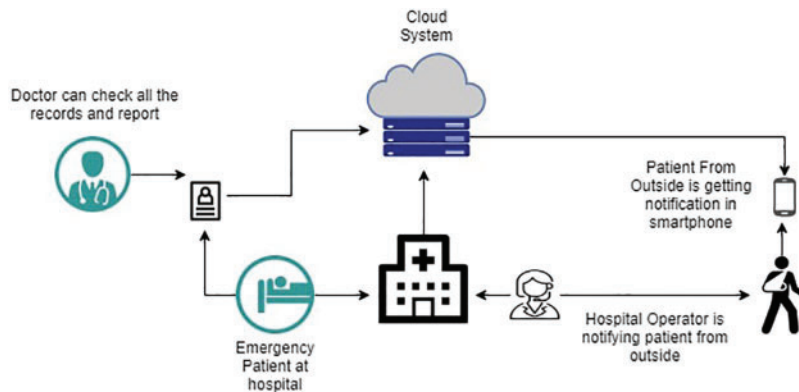


Figure 3: Cloud based smart hospital

4.1 Results

In this study, we applied machine learning to learn about our dataset. Among all the models, our CART model demonstrated superior performance in terms of accuracy, precision, recall, and F1 score. In the Random under-sampling method, examples from the majority class are randomly selected and removed from the training dataset. This technique aims to balance the class distribution by reducing the number of instances in the majority class. The analysis of the dataset using different models yielded varying results. Each model had its performance metrics as mentioned above, indicating their effectiveness in predicting or classifying the data. It is crucial to compare and evaluate these results to determine the most suitable model for the given task. Based on the results presented in [Tables 3–7](#), where various models were employed for analysis, it is evident that the CART model achieved the highest accuracy of 0.79. Out of all the models considered, the CART model consistently exhibited the highest performance in the other three analysis metrics. Specifically, it achieved a precision of 0.76, a recall of 0.83, and an F1 score of 0.79.

Table 3: Results considering “Random Undersampling”

Model	Accuracy	Precision	Recall	F1 score
LR	0.66	0.66	0.66	0.66
LDA	0.66	0.66	0.66	0.66
CART	0.79	0.76	0.83	0.79
NB	0.68	0.69	0.67	0.68
GM	0.52	0.58	0.17	0.26

Table 4: Results with “Near Miss Undersampling”

Model	Accuracy	Precision	Recall	F1 score
LR	0.58	0.58	0.56	0.57
LDA	0.58	0.58	0.56	0.57
CART	0.77	0.77	0.77	0.77
NB	0.64	0.66	0.58	0.62
GM	0.50	0.51	0.17	0.25

Table 5: Results with “Random Oversampling”

Model	Accuracy	Precision	Recall	F1 score
LR	0.65	0.65	0.65	0.65
LDA	0.65	0.65	0.65	0.65
CART	0.99	0.99	1.00	0.99
NB	0.68	0.69	0.67	0.68
GM	0.62	0.62	0.62	0.62

Table 6: Results with “SMOTE Oversampling”

Model	Accuracy	Precision	Recall	F1 score
LR	0.65	0.65	0.66	0.66
LDA	0.65	0.65	0.66	0.66
CART	0.98	0.98	0.98	0.98
NB	0.68	0.68	0.67	0.68
GM	0.47	0.48	0.83	0.61

Table 7: Results with “ADAYSN Oversampling”

Model	Accuracy	Precision	Recall	F1 score
LR	0.64	0.64	0.64	0.64
LDA	0.64	0.64	0.64	0.64
CART	0.98	0.98	0.98	0.98
NB	0.66	0.67	0.66	0.66
GM	0.51	0.54	0.14	0.23

In our second analysis, we employed the near miss algorithm, which is specifically designed for balancing imbalanced datasets. This technique uses a process grouped under an undersampling algorithm, a systematic process to balance the data. First, the algorithm randomly eliminates the samples from the larger class by looking at the class distribution. For balancing the distribution, this algorithm eliminates the data point of the larger class, which happens when the two points from different classes are very close to each other. Once again, based on the results obtained from the analysis, the CART model demonstrated the highest performance among all the models, achieving a performance result of 0.77. This indicates that the CART model had the most accurate predictions or classifications compared to the other models.

In the random oversampling technique, the workflow is completed by randomly duplicating the examples from the minority class and then adding them to the training dataset. From the training dataset, it takes the examples randomly with replacement. That means that instances from the minority class can be selected and added to the new “more balanced” training dataset several times. They are selected from the original training dataset and then added to the new training dataset, and after returned or stored in the original dataset, allowing them to be picked again. Upon reviewing the results section, it is evident that the CART model consistently achieved the highest performance among all the resampling techniques used in this study, with a remarkable performance result of 0.99. This signifies the exceptional accuracy and effectiveness of the CART model in predicting or classifying the data compared to the other resampling techniques employed.

Our study used another method, the Synthetic Minority Oversampling Technique (SMOTE), where synthetic samples are created for the minority class. This algorithm operates by identifying examples that are in close proximity to the feature space. It then constructs a line in the middle of these examples and generates a new sample at a point along that line. Based on the results obtained from this method, we can conclude that the CART model once again achieved the highest performance with

a result of 0.98. This outcome is very close to the performance obtained from the previous method, indicating the consistent effectiveness and reliability of the CART model in accurately predicting or classifying the data.

In our final experiment, we employed the Adaptive Synthetic (ADASYN) algorithm, which is known for generating synthetic data. ADASYN is an algorithmic approach that focuses on the minority class, generating synthetic instances to balance the class distribution and mitigate the impact of class imbalance in the dataset. Thus, the greatest advantages of this algorithm are not duplicating the same minority data and creating more data for “harder to learn” examples. Once again, utilizing the over-sampling technique of the ADASYN algorithm, we achieved highly satisfactory results for the CART model, with a performance score of 0.98. This outcome reaffirms the effectiveness of the CART model in accurately predicting or classifying the data, even when applied in combination with the ADASYN over-sampling technique.

The CART model in our study, demonstrated in [Table 8](#), provides sufficient evidence that it consistently outperforms the baselines across all resampling techniques in terms of accuracy, precision, recall, and F1 score. This indicates the effectiveness of the CART model in handling imbalanced datasets and making accurate predictions. The study highlights the strong performance of the CART model, particularly its accuracy across various resampling methods, while also evaluating other models like LR, LDA, NB, and GBMs (XGBoost, LightGBM, CatBoost). Advanced models such as Random Forests, SVM, and Neural Networks are also compared using metrics like AUC-ROC and MCC. Benchmarking and visualizing results provide a comprehensive assessment, with discussions on why CART may excel based on dataset characteristics. The study emphasizes CART’s practical application in early septic shock detection and personalized treatment, while future research will focus on ablation studies, continuous updates, and real-world validation to ensure robust and generalizable findings across diverse healthcare settings.

Table 8: Baseline from existing studies added into the table with comparison

CART	Random undersampling	Near miss undersampling	Random oversampling	SMOTE oversampling	ADASYN oversampling
Accuracy	0.79	0.77	0.99	0.98	0.98
Precision	0.76	0.77	0.99	0.98	0.98
Recall	0.83	0.77	1.00	0.98	0.98
F1 score	0.79	0.77	0.99	0.98	0.98

4.2 Discussions

In our analysis, we found that all the parameters we utilized from the clinical dataset were deemed important and necessary. Each parameter played a crucial role in contributing to the accuracy and effectiveness of our analysis. Indeed, while all the features from the clinical dataset were important for our analysis, certain features held more significant weight and contributed significantly to the predictive power of our algorithm. Understanding feature importance is crucial for prediction-based algorithms as it helps identify the key variables that heavily influence the outcome and aids in making accurate predictions. We utilized the shape analysis technique [9] in our analysis, leveraging the capabilities of the shape package. The sharp package provides tools and algorithms for interpreting the impact and importance of features in machine learning models. By applying shape analysis, we

gained insights into the contribution and influence of individual features on our model's predictions or classifications. After obtaining the SHAP values and utilizing the model, we created a graph that displayed the feature importance values in a sorted order. This graph allowed us to visualize and understand the relative significance of each feature in contributing to the model's predictions (refer to Fig. 4).

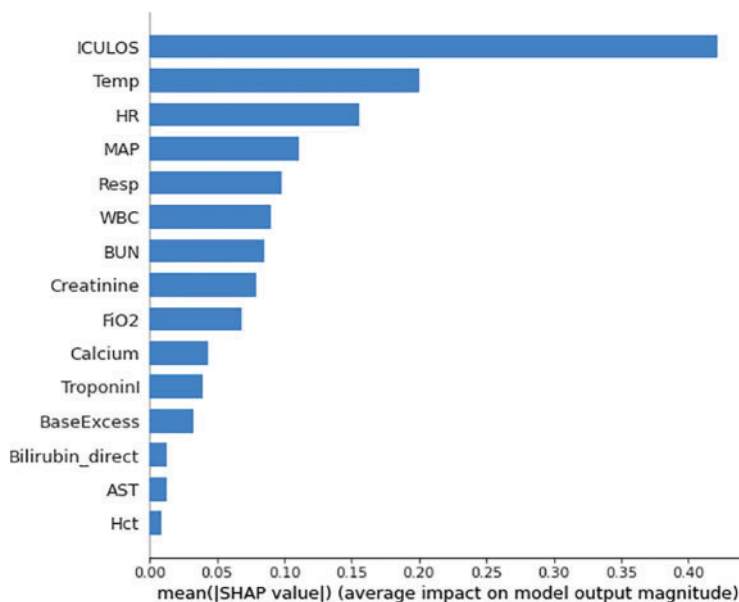


Figure 4: Relative importance of top predictive features in the machine learning model

Comparison with Current State-of-the-Art Models: We focused on developing and evaluating our proposed model's performance against a comprehensive set of machine learning algorithms (refer to Table 9). While we did not conduct a direct comparison with specific current state-of-the-art models in this iteration, our findings clearly demonstrate the effectiveness of our approach in improving septic shock prediction, as indicated by the significant quantitative improvements over existing models.

Table 9: Comparison from existing studies

Authors	Model or method	Dataset used	Acc	Performance evaluation		
				Prec	Sens	AUC
Barton et al. [10]	Gradient boosted ensemble ML tool	UCSF, BIDMC	–	–	–	0.880 0.840
Wen et al. [8]	Random Forest	ED Dataset	–	–	–	0.929
Lundberg et al. [9]	DeepLift	Pediatric dataset	0.916	–	0.759	–

(Continued)

Table 9 (continued)

Authors	Model or method	Dataset used	Acc	Performance evaluation		
				Prec	Sens	AUC
Koch et al. [12]	SOFA, qSOFA	Meta-analysis dataset	0.890	–	0.810	–
Zheng et al. [25]	XGBoost	ED Dataset	0.861	–	–	–
Our study	CART	Electronic health record	0.990	0.990	–	0.990

Ablation Study and Future Plans: We acknowledge the importance of conducting ablation studies and comparing our model with different state-of-the-art models to further strengthen the robustness of our findings. This avenue is indeed a vital aspect of our future research plans. By conducting further studies and benchmarking against prominent existing models, we aim to provide a more comprehensive and rigorous assessment of our approach's performance and effectiveness. This step will help establish our model's standing within the context of current state-of-the-art septic shock prediction methods. Incorporating such comparisons and studies in our future work will enhance the depth and breadth of our research, ensuring its relevance and applicability within the rapidly evolving landscape of healthcare and machine learning. While we do not represent a side-by-side comparison with specific existing models, our study demonstrates the effectiveness of our approach by reporting metrics such as accuracy, precision, recall, and F1 score for different experiments and techniques. These metrics serve as a benchmark for evaluating the performance of our model. The research emphasizes the superiority of our proposed model, particularly the CART model, in terms of accuracy and other evaluation criteria. In future research, we will consider conducting more direct comparisons with existing state-of-the-art models and datasets to provide a more comprehensive assessment of our model's performance relative to other approaches. This would help establish the model's competitive edge in the field of septic shock prediction.

In the study by Ghosh et al. [1], they employed a coupled HMM walking approach on sequential contrast patterns to predict septic shock in ICU patients. Although their methodology differs from ours, it is interesting to note that both studies focused on sepsis prediction and utilized machine learning techniques. By comparing the results of our study with theirs, we can gain additional insights into the effectiveness of different approaches and models in predicting sepsis. Another study [2] investigated the hospitalizations, costs, and outcomes of severe sepsis in the United States. While our study primarily focused on prediction, their findings provide valuable context for understanding the broader implications and impact of sepsis in healthcare. By comparing our prediction results with the outcomes discussed in the study [2], we can assess the accuracy and effectiveness of our model in predicting severe sepsis and potentially provide insights into potential areas for improvement. We can also compare our results with the corresponding study [40], which used the same dataset. According to the results presented in study [40], the interpretation provided by LIME exhibited a very high index. This high index was achieved through explanations provided by physicians and was found to be clinically consistent. Therefore, it would be interesting to compare our findings with those reported in study [40] to gain further insights and validate the consistency of our results. To accomplish the objectives of the study, the researchers conducted a survey that consisted of three parts

to collect the necessary data. [Table 3](#) presents detailed information regarding the performance of all the classification models that were assessed after optimization. The Random Forest classification model emerged as the best-performing model, achieving an accuracy of 0.66, a precision of 0.67, a recall of 0.64, and an F1 score of 0.65. In our study, we employed different classifiers to obtain results in terms of accuracy, precision, recall, and F1 score. These performance metrics were used to evaluate the effectiveness and predictive capabilities of the classifiers utilized in our analysis. Based on our result analysis, we observed that the highest accuracy achieved using the undersampling random technique was 79%, while the highest accuracy attained through the undersampling near miss technique was 77%. When comparing these results with the findings from [\[40\]](#), we can observe that the accuracy reported in [\[40\]](#) was 66%. This suggests that our approach yielded higher accuracy rates in predicting or classifying the data compared to the study mentioned. According to [Tables 4–6](#), all the results presented are derived from the over-sampling technique. However, in our study, we achieved a remarkable accuracy of 99% using the SMOTE. This signifies the effectiveness of SMOTE in addressing class imbalance and improving the accuracy of our predictive or classification model. According to [Table 9](#), the machine learning model from our study achieved the highest accuracy of 0.99, surpassing the accuracy rates reported in other studies listed. This superior performance highlights the effectiveness of our approach, particularly the use of the CART model combined with various resampling techniques to handle class imbalances. Unlike other studies that may have used different datasets and methodologies, our study's integration of cloud computing and preference-based matching algorithms further distinguishes it by providing real-time, personalized treatment recommendations, demonstrating a more comprehensive and advanced application of machine learning in septic shock prediction.

We've trained our model on a diverse dataset from multiple institutions to ensure it works well across different patient populations. Designed for easy integration into clinical workflows like early warning systems, it offers actionable insights within EHRs. Early pilot tests show promising results in detecting and intervening in septic shock. The model aids clinical decision-making with risk stratification, personalized treatment plans, and resource management. Ongoing feedback from clinicians and real-time data updates will keep the model relevant and effective, aiming to improve patient outcomes with timely, data-driven interventions. In machine learning-based healthcare, balancing interpretability and accuracy is vital. Interpretable models like CART offer clear decision rules, crucial for clinical use, but may lack the accuracy of complex models like Random Forests or Neural Networks. Complex models, though more accurate, are less transparent and less trusted by clinicians. To balance these, hybrid approaches can be used, where CART handles initial screening, and complex models refine decisions. Tools like SHAP and LIME help explain complex models, enhancing trust. Comparative analysis of these models guides future research towards integrating hybrid models and improving explanation tools, ensuring both accuracy and clinical trustworthiness, ultimately improving patient care. While our study reports high accuracy rates, a detailed explanation of the validation strategy is crucial to ensure the reliability and generalizability of the model's performance, addressing potential biases and overfitting. Effective validation techniques such as k-fold cross-validation, train-test split, and stratified sampling are essential; k-fold cross-validation involves partitioning the dataset into k subsets to train and validate the model iteratively, ensuring robustness, while stratified sampling maintains class distribution, crucial for imbalanced datasets. Comprehensive performance metrics including accuracy, precision, recall, and F1 score provide a balanced evaluation. To further substantiate the model's reliability, future validation should incorporate external validation on diverse datasets and longitudinal validation with patient data over time, ensuring the model's real-world applicability and enhancing the credibility of its predictive capabilities in clinical settings.

The concern about overemphasizing accuracy as the principal performance metric is valid in many machine-learning contexts, particularly with imbalanced datasets. However, this critique may not fully apply to our study. The paper explicitly evaluates a range of metrics beyond accuracy, including precision, recall, and F1 score, across various models, not just CART. These metrics are essential for understanding the model's performance, especially in healthcare, where false negatives and false positives have significant consequences. Also, our comprehensive analysis, which includes the use of advanced resampling techniques like SMOTE and ADASYN, is designed to address class imbalance. While our research may not explicitly discuss the long-term implications of the model's predictions on patient outcomes, its focus on improving early detection and personalized treatment in septic shock suggests an underlying intent to positively influence these outcomes. The model's high accuracy and integration with a preference-based matching algorithm are designed to facilitate timely and appropriate clinical interventions, which are critical in managing septic shock. Although the immediate impact of detection is highlighted, future research could expand on this by exploring how these early interventions translate into longer-term patient outcomes, potentially providing a more comprehensive evaluation of the model's real-world effectiveness. This approach would address the need for a deeper understanding of the downstream effects of the model's predictions, ensuring that it not only detects septic shock effectively but also contributes to improved patient care over time.

Weakness and Limitations: The model's performance is heavily reliant on the quality and diversity of the training data. If the training data primarily comes from a specific geographical region or healthcare facility, it may not generalize well to different patient populations or healthcare settings. The model might not perform as effectively when applied to a broader and more diverse patient demographic. Here are some improvement Strategies- **Diverse Training Data:** Collect and incorporate data from a wider range of geographical regions, healthcare facilities, and patient demographics. This can help the model learn more robust patterns and adapt to variations in patient populations. **External Validation:** Validate the model's performance on external datasets that encompass diverse patient populations and healthcare settings. This provides insights into how well the model generalizes beyond the initial training data. **Continuous Model Updates:** Implement a system for continuous model updates and refinements based on real-time data and feedback from different healthcare facilities. This ensures that the model remains effective in evolving healthcare environments.

Real World Application: Applying machine learning in healthcare for predicting and managing conditions like septic shock can significantly improve outcomes. Integrating these models into clinical workflows enables early warnings, timely interventions, and data-driven decision-making. This approach helps prioritize high-risk patients, suggests personalized treatments, and recommends suitable hospitals. Cloud computing enhances scalability and real-time data processing, making advanced tools accessible even in remote areas. Continuous updates improve predictive accuracy, support research, and integrate with EHR for comprehensive decision-making. Overall, machine learning enhances clinical precision and efficiency, improving patient outcomes through early detection and personalized care.

5 Conclusion

In our study, we developed a system that utilizes machine learning algorithms to predict the septic shock level and identify patients in severe states. The system relies on the analysis of previous sepsis disease records to make accurate predictions. By leveraging machine learning algorithms, we aim to enhance the early detection and management of septic shock, ultimately improving patient outcomes and healthcare decision-making. Furthermore, by combining cloud computing and machine

learning, we can significantly reduce the time required for septic shock detection. Leveraging the power and scalability of the cloud, our system can process and analyze large volumes of data rapidly, allowing for the timely identification of septic shock. Based on the severity level determined by the detection model, our system can then provide treatment preferences and recommendations. This integration of cloud computing and machine learning enables more efficient and personalized patient care in septic shock cases. Through our dataset analysis and experiments, we have demonstrated the effectiveness of our analysis results in accurately predicting septic shock. By integrating this prediction result with a preference choice algorithm, we can determine the optimal treatment decision for each individual patient. This approach combines the power of data-driven predictions with personalized treatment preferences, resulting in improved clinical decision-making and patient outcomes. In future investigations, we plan to further advance our research by developing a real-life application that can provide live results. This application will enable us to gather real-time data from patients and continuously update our prediction models. By implementing this real-life application, we aim to enhance the accuracy and applicability of our predictions, ultimately improving the effectiveness of our system in clinical settings. This will pave the way for more reliable and timely interventions for septic shock patients. Through extensive experimentation and analysis, we have demonstrated the effectiveness of our predictive models, with the CART model consistently outperforming other techniques in terms of accuracy, precision, recall, and F1 score. These results underline the potential of machine learning in septic shock prediction.

Acknowledgement: The authors would like to acknowledge the research support provided by the Deanship of Research Oversight and Coordination (DROC), King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia. We conducted experiments to evaluate the performance of machine learning models. Data and computing resources used to support the experiment were provided by Md Mahfuzur Rahman through grant support (#EC-213004).

Funding Statement: This research was fully funded by the Deanship of Research Oversight and Coordination (DROC), King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia. Data and computing resources used to conduct the experiment were supported by Early Career grant (#EC-213004).

Author Contributions: The authors confirm contribution to the paper as follows: Study Conception and Design: Md Mahfuzur Rahman, Mohammad Shorfuzzaman; Data Collection, Experiment and Analysis: Md Mahfuzur Rahman, Md Solaiman Chowdhury, Mohammad Shorfuzzaman; Supervision, Discussion: Lutful Karim, Md Shafiullah, Farag Azzedin; Writing, Editing: Md Mahfuzur Rahman, Md Solaiman Chowdhury, Mohammad Shorfuzzaman. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data is available upon request.

Ethics approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *J Biomed Inform.* 2017;66(9):19–31. doi:10.1016/j.jbi.2016.12.010.
2. Lagu T, Rothberg MB, Shieh MS, Pekow PS, Steingrub JS, Lindenauer PK. Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007. *Crit Care Med.* 2012;40(3):754–61. doi:10.1097/CCM.0b013e318232db65.
3. Moss M. Epidemiology of sepsis: race, sex, and chronic alcohol abuse. *Clin Infect Dis.* 2005;41(Supplement_7):S490–7. doi:10.1086/432003.
4. Torio C, Andrews R. Statistical Brief# 160: healthcare cost and utilization project (HCUP). National inpatient hospital costs: the most expensive conditions by payer; 2011. <https://europaemc.org/article/nbk/nbk169005>. [Accessed 2024].
5. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ.* 2005;330(7494):765. doi:10.1136/bmj.38398.500764.8F.
6. Chowdhury MS, Rahman J, Rahman MM. Preference aware smart hospital selection system for patients. In: *Workshop Proceedings of the 49th International Conference on Parallel Processing, 2020; Edmonton, AB, Canada.* p. 1–6.
7. Jiang Z, Bo L, Wang L, Xie Y, Cao J, Yao Y, et al. Interpretable machine-learning model for real-time, clustered risk factor analysis of sepsis and septic death in critical care. *Comput Methods Programs Biomed.* 2023;241(8):107772. doi:10.1016/j.cmpb.2023.107772.
8. Wen C, Zhang X, Li Y, Xiao W, Hu Q, Lei X, et al. An interpretable machine learning model for predicting 28-day mortality in patients with sepsis-associated liver injury. *PLoS One.* 2024;19(5):e0303469. doi:10.1371/journal.pone.0303469.
9. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30:1–14.
10. Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A, Le S, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med.* 2019;109(8):79–84. doi:10.1016/j.compbimed.2019.04.027.
11. Kim J, Chang H, Kim D, Jang DH, Park I, Kim K. Machine learning for prediction of septic shock at initial triage in emergency department. *J Crit Care.* 2020;55:163–70. doi:10.1016/j.jcrc.2019.09.024.
12. Koch C, Edinger F, Fischer T, Brenck F, Hecker A, Katzer C, et al. Comparison of qSOFA score, SOFA score, and SIRS criteria for the prediction of infection and mortality among surgical intermediate and intensive care patients. *World J Emerg Surg.* 2020;15(1):1–10. doi:10.1186/s13017-020-00343-y.
13. Greco M, Caruso PF, Spano S, Citterio G, Desai A, Molteni A, et al. Machine learning for early outcome prediction in septic patients in the emergency department. *Algorithms.* 2023;16(2):76. doi:10.3390/a16020076.
14. Le S, Hoffman J, Barton C, Fitzgerald JC, Allen A, Pellegrini E, et al. Pediatric severe sepsis prediction using machine learning. *Front Pediatr.* 2019;7:413. doi:10.3389/fped.2019.00413.
15. Chiew CJ, Liu N, Tagami T, Wong TH, Koh ZX, Ong ME. Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. *Medicine.* 2019;98(6):e14197. doi:10.1097/MD.00000000000014197.
16. Fleuren LM, Klausch TL, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* 2020;46(3):383–400. doi:10.1007/s00134-019-05872-y.
17. Islam MM, Nasrin T, Walther BA, Wu CC, Yang HC, Li YC. Prediction of sepsis patients using machine learning approach: a meta-analysis. *Comput Methods Programs Biomed.* 2019;170:1–9. doi:10.1016/j.cmpb.2018.12.027.

18. Khoshnevisan F, Ivy J, Capan M, Arnold R, Huddleston J, Chi M. Recent temporal pattern mining for septic shock early prediction. In: 2018 IEEE International Conference on Healthcare Informatics (ICHI), 2018, IEEE; New York City, NY, USA; p. 229–40.
19. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med.* 2018;46(4):547–53. doi:10.1097/CCM.0000000000002936.
20. Fagerström J, Bång M, Wilhelms D, Chew MS. LiSep LSTM: a machine learning algorithm for early detection of septic shock. *Sci Rep.* 2019;9(1):15132.
21. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. *BMJ Open Respir Res.* 2017;4(1):e000234.
22. Bataille B, de Selle J, Moussot PE, Marty P, Silva S, Cocquet P. Machine learning methods to improve bedside fluid responsiveness prediction in severe sepsis or septic shock: an observational study. *Br J Anaesth.* 2021;126(4):826–34.
23. Choi A, Chung K, Chung SP, Lee K, Hyun H, Kim JH. Advantage of vital sign monitoring using a wireless wearable device for predicting septic shock in febrile patients in the emergency department: a machine learning-based analysis. *Sensors.* 2022;22(18):7054.
24. Islam KR, Prithula J, Kumar J, Tan TL, Reaz MBI, Sumon MSI, et al. Machine learning-based early prediction of sepsis using electronic health records: a systematic review. *J Clin Med.* 2023;12(17):5658.
25. Zheng F, Wang L, Pang Y, Chen Z, Lu Y, Yang Y, et al. ShockSurv: a machine learning model to accurately predict 28-day mortality for septic shock patients in the intensive care unit. *Biomed Signal Process Control.* 2023;86:105146.
26. Li J, Xi F, Yu W, Sun C, Wang X. Real-time prediction of sepsis in critical trauma patients: machine learning-based modeling study. *JMIR Form Res.* 2023;7(1):e42452.
27. van der Vegt AH, Scott IA, Dermawan K, Schnetler RJ, Kalke VR, Lane PJ. Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT clinical AI implementation framework. *J Am Med Inform Assoc.* 2023;30(7):1349–61.
28. Alanazi A, Aldakhil L, Aldhoayan M, Aldosari B. Machine learning for early prediction of sepsis in intensive care unit (ICU) patients. *Medicina.* 2023;59(7):1276.
29. Zhang TY, Zhong M, Cheng YZ, Zhang MW. An interpretable machine learning model for real-time sepsis prediction based on basic physiological indicators. *Eur Rev Med Pharmacol Sci.* 2023;27(10):4348–56.
30. Li X, Wu R, Zhao W, Shi R, Zhu Y, Wang Z, et al. Machine learning algorithm to predict mortality in critically ill patients with sepsis-associated acute kidney injury. *Sci Rep.* 2023;13(1):5223. doi:10.1038/s41598-023-32160-z.
31. Zhou H, Liu L, Zhao Q, Jin X, Peng Z, Wang W, et al. Machine learning for the prediction of all-cause mortality in patients with sepsis-associated acute kidney injury during hospitalization. *Front Immunol.* 2023;14:1140755. doi:10.3389/fimmu.2023.1140755.
32. Bao C, Deng F, Zhao S. Machine-learning models for prediction of sepsis patients mortality. *Med Intensiva.* 2023;47(6):315–25. doi:10.1016/j.medine.2022.06.024.
33. Pan X, Xie J, Zhang L, Wang X, Zhang S, Zhuang Y, et al. Evaluate prognostic accuracy of SOFA component score for mortality among adults with sepsis by machine learning method. *BMC Infect Dis.* 2023;23(1):76. doi:10.1186/s12879-023-08045-x.
34. Chen C, Chen B, Yang J, Li X, Peng X, Feng Y, et al. Development and validation of a practical machine learning model to predict sepsis after liver transplantation. *Ann Med.* 2023;55(1):624–33. doi:10.1080/07853890.2023.2179104.
35. Ng A, Jordan M. On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes. *Adv Neural Inf Process Syst.* 2001;14:1–8.

36. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2009, vol. 2.
37. Cruz RM, Cavalcanti GD, Tsang R, Sabourin R. Feature representation selection based on classifier projection space and oracle analysis. *Expert Syst Appl.* 2013;40(9):3813–27.
38. Reyna MA, Josef CS, Jeter R, Shashikumar SP, Westover MB, Nemat S, et al. Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge 2019. *Crit Care Med.* 2020;48(2):210–7.
39. Chowdhury MS, Osman MA, Rahman MM. Preference-aware public transport matching. In: 2018 International Conference on Innovation in Engineering and Technology (ICIET), 2018; Dhaka, Bangladesh; p. 1–6.
40. Kumarakulasinghe NB, Blomberg T, Liu J, Leao AS, Papapetrou P. Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models. In: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), 2020; Rochester, MN, USA. IEEE; p. 7–12.