International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⪀ Peer-reviewed & Refereed journal ⪀ Vol. 14, Issue 5, May 2026
DOI: 10.17148/IJIREEICE.2026.14542

AN INTELLIGENT SEPSIS PREDICTION AND PATIENT MONITORING SYSTEM USING MACHINE LEARNING

Dr. T. Amalraj Victoire[1], V. Swetha[2]

Professor, Department of Master Computer Application, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605 107, India[1]

PG Student, Department of MCA, Sri Manakula Vinayagar Engineering College (Autonomous), Madagadipet, Puducherry – 605 107, India[2]

Abstract: In critical care settings, early detection of medical emergencies plays a crucial role. Traditional methods often rely on healthcare professionals manually observing patients, which can lead to the oversight of early signs of declining health. This study introduces a real-time monitoring system that utilizes machine learning to analyze essential physiological data and identify potential health risks before they deteriorate. The system focuses on patients with sepsis and monitors key vital signs, including heart rate, oxygen saturation, body temperature, respiratory rate, systolic blood pressure, and diastolic blood pressure. We employed various machine learning algorithms, including Logistic Regression, Decision Tree, and Random Forest, to effectively assess the risk of sepsis. Careful data preprocessing and feature selection significantly enhanced the performance of the models. Among the models tested, the Random Forest classifier achieved the highest accuracy. Additionally, we integrated the trained model into an easy-to-use dashboard built on Streamlit, enabling real-time patient monitoring, anomaly detection, and comprehensive risk analysis. This system is designed to assist healthcare professionals in promptly addressing hidden medical emergencies, thereby improving clinical decision-making. Furthermore, SHapley Additive Explanations (SHAP) analysis was used to explain the model predictions, enhancing transparency and trust.

Keywords: Machine Learning, Sepsis Prediction, Healthcare Monitoring, Random Forest, Data Analytics, Streamlit Dashboard

I. INTRODUCTION

Sepsis is a life-threatening condition characterized by acute organ dysfunction caused by an infection-related inflammatory response. It has become one of the leading causes of death globally, especially in intensive care units (ICUs) [1], [2]. It is also a significant challenge in ICUs due to its rapid progression and swift deterioration of patient conditions. Early detection significantly reduces mortality risk. [3].

Traditionally, clinical methods depend on periodic checks and manual patient assessments. However, recognizing subtle early warning signs is complex due to the intricate and dynamic nature of the disease. There is a growing demand for intelligent systems that can continuously monitor patient data and provide early predictions. [4].

Machine learning has emerged as a powerful tool in healthcare, predictive analytics, and decision-making. By examining large volumes of patient data, machine learning models can uncover hidden patterns and relationships that are not readily apparent with conventional methods. [5], [6]. As a result, machine learning is well-suited for early detection of critical conditions, including sepsis.

In this study, we present a real-time sepsis prediction system based on multiple machine learning algorithms. The system utilizes vital signs, including heart rate, oxygen saturation, temperature, respiratory rate, systolic blood pressure, and diastolic blood pressure, to predict sepsis. The Random Forest algorithm was implemented to enhance prediction accuracy. [7].

The proposed system features an interactive dashboard built on Streamlit offering real-time visualization of patient data and prediction results for monitoring within the ICU. The alert function is activated when a patient's condition reaches a critical level, allowing healthcare professionals to respond promptly to changes. [8].

DOI: 10.17148/IJIREEICE.2026.14542

The primary objective of the proposed system is to develop an efficient, accurate, and user-friendly tool for early sepsis prediction and real-time monitoring. It aims to support healthcare professionals in making rapid and informed decisions that improve patient outcomes.

II. RELATED WORK

Several studies have utilized machine learning techniques to monitor healthcare data and predict diseases, particularly critical illnesses such as sepsis. Clinical scoring systems and manual observations are the most commonly used standard practices in hospitals. However, early and accurate detection is rarely achieved due to their dependence on preset thresholds and human intervention [9], [10].

Several machine learning models, including Logistic Regression and Decision Trees, have been applied in recent years to predict medical conditions based on patient data. These are simple and transparent techniques; however, they may fail in more complex and nonlinear situations typically encountered in healthcare datasets [11].

This has led to improved prediction accuracy using Random Forest and other ensemble learning techniques. Random Forest, in particular, has the ability to handle high-dimensional data and complex relationships between features efficiently [12], [13].

Table 1: Comparison of Existing Machine Learning Models used in healthcare prediction

Study	Method	Accuracy	Limitation
Kumar et al.	Logistic Regression	85%	Low accuracy
Singh et al.	Decision Tree	88%	Overfitting
Chen et al.	Random Forest	92%	No explainability

Other studies have developed systems for real-time patient monitoring with web-based dashboards, allowing healthcare personnel to continuously monitor patient status and receive alerts for abnormal pattern detection [14].

However, most existing solutions have limitations, such as the absence of real-time prediction, limited user-friendly interfaces, and a lack of integrated alerting mechanisms [15]. Therefore, a comprehensive system is required to integrate accurate machine learning prediction with real-time monitoring and visualization.

The proposed system addresses this gap by integrating a machine learning model with an interactive dashboard and a real-time alert system, providing an all-in-one solution for early sepsis prediction and patient monitoring.

III. PROPOSED SYSTEM

The system aims to deliver an intelligent solution for the early detection of sepsis and real-time monitoring of patient conditions. By combining machine learning approaches with an interactive dashboard, the system analyzes patients' vital signs and produces indications of sepsis risk. This allows doctors to identify potential cases at an early stage and respond with appropriate treatment.

The system requires input features such as heart rate, oxygen saturation, body temperature, respiration rate, systolic BP, and diastolic BP. These features are provided to a trained machine learning model for sepsis prediction. The model is trained using various algorithms, and the Random Forest classifier is selected due to its high accuracy and efficiency in handling medical data [16], [17].

Table 2. Summary of System Modules

Module	Function
Input Module	Collect patient data
Preprocessing	Clean & normalize data
ML Module	Predict sepsis
Monitoring	Display real-time data
Explainable AI	Interpret predictions

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14542

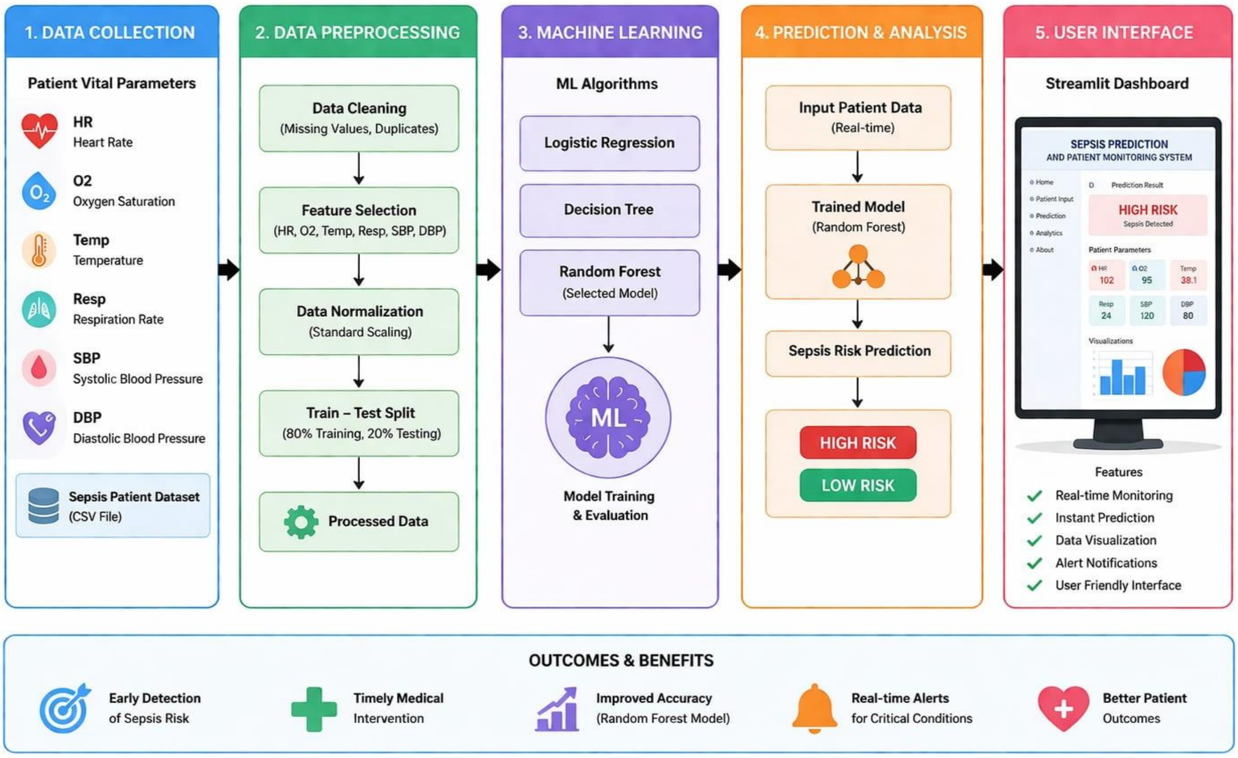

Figure 1. Architecture of the proposed sepsis prediction and ICU monitoring system

The proposed system architecture consists of several steps: initial data input, preprocessing phase, model prediction, and visualization. The data is entered manually by the user or imported from a dataset and then preprocessed to be compatible with the model. For prediction, the model receives data in the correct format and outputs a prediction score to estimate the risk.

The prediction result is visualized to the user through a Streamlit-powered dashboard, which provides a clear view of the patient's status in real time. The user can observe risk indicators, graphs, and patient parameters through a simple interface. Alert systems are included to notify the doctor if the values exceed predefined thresholds. Additionally, the system records each patient's data and prediction results over time, providing a simple and practical analysis framework.

IV. DATASET DESCRIPTION

The dataset for this project was obtained from the PhysioNet Sepsis Challenge. It contains real-world ICU patient records and vital health parameters that help predict sepsis risk. There are approximately 546,123 patient records with 44 clinical features—each row represents a patient observation, and each column corresponds to a specific physiological parameter. The features selected for this study include heart rate (HR), oxygen saturation (O2Sat), body temperature (Temp), respiration rate (Resp), systolic blood pressure (SBP), and diastolic blood pressure (DBP). These metrics are commonly used in clinical monitoring and play a significant role in identifying early signs of infection and organ dysfunction related to sepsis. The target variable is the SepsisLabel, which indicates whether a patient is at risk of sepsis.

It is important to understand the distribution and behavior of these features. Some attributes contain missing values or outliers, which can adversely affect the performance of machine learning models. Therefore, preprocessing techniques were applied to ensure the quality and consistency of the data [18], [19].

Prior to model development, preprocessing techniques such as handling missing values, correcting inconsistencies, and normalizing features were applied to enhance data quality and model performance. The dataset was split into training and testing sets in an 80:20 ratio, which helps ensure proper model validation and generalization on unseen patient data.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026
DOI: 10.17148/IJIREEICE.2026.14542

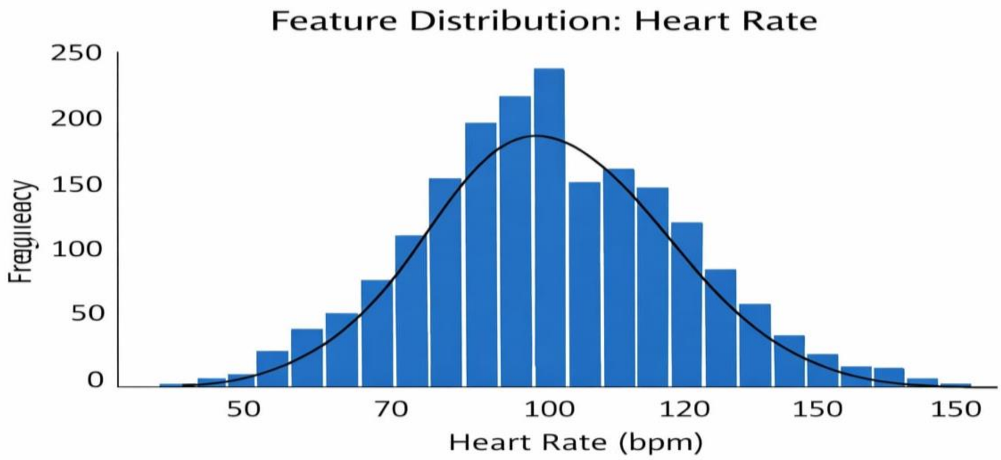

Figure 2. Feature Distribution: Heart Rate

Table 3. Dataset Features and Normal Ranges

Feature	Description	Normal Range
Heart Rate	Beats per minute	60 - 100bpm
Temperature	Body temperature	36.5 - 37.5 °C
Oxygen Saturation	Blood oxygen level	95 – 100%
Respiration Rate	Breaths per minute	12 – 20
Systolic BP	Blood pressure (upper)	90 –120 mmHg
Diastolic BP	Blood pressure (lower)	60 – 80 mmHg

V. METHODOLOGY

The proposed system's methodology consists of a series of steps ranging from data preprocessing to model prediction and evaluation. Each phase was carefully designed to ensure that the data is adequately prepared and that the machine learning model operates effectively.

First, the dataset was preprocessed, during which missing values were handled, and inconsistencies were removed. This crucial step ensured that the dataset was clean and ready for training. Additionally, feature scaling was applied to normalize the values, thereby improving the machine learning model's performance.

Following the preprocessing stage, the dataset was partitioned into training and testing sets. The training set was used to develop the model, whereas the testing set was used to assess its performance. This division is essential for validating the model's effectiveness on data that it has not encountered before.

This study explored various machine learning algorithms, specifically Logistic Regression, Decision Tree, and Random Forest. Each model was trained on the training dataset and evaluated for accuracy. Ultimately, the Random Forest model was selected as the preferred model due to its superior accuracy and its ability to handle complex data patterns [20].

After training, the model was integrated into the system for real-time prediction. User-provided input data undergoes preprocessing before being fed into the trained model. The model then estimates the probability of sepsis risk, which is subsequently presented on the dashboard.

To ensure the model's reliability in practical healthcare settings, the system continuously evaluates its performance using standard metrics such as accuracy. This process ensures that the predictions made by the model are reliable and applicable in real-world scenarios.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14542

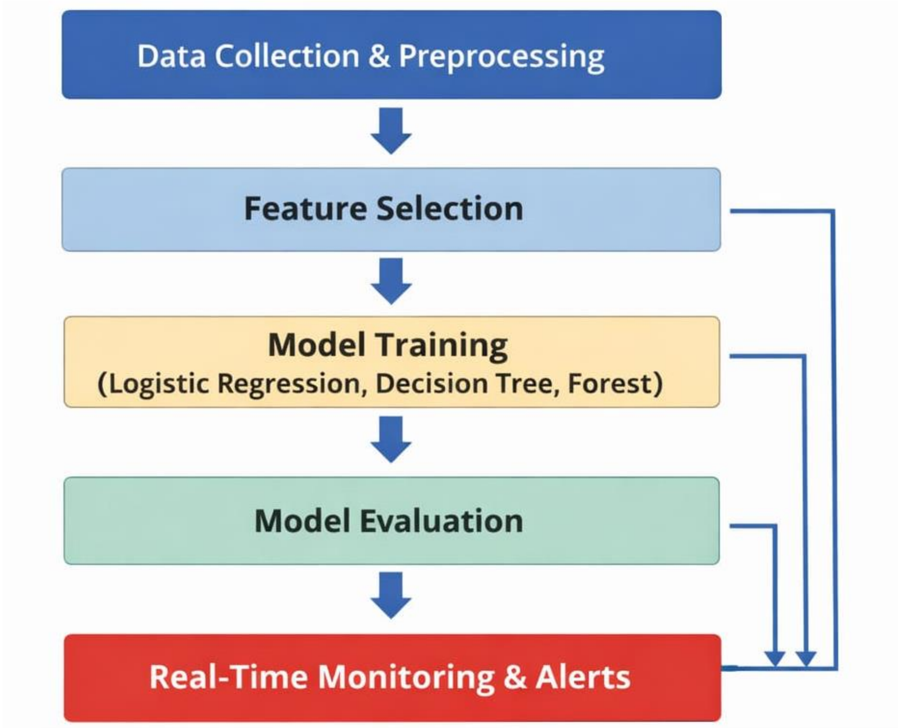

Figure 3. Machine Learning Workflow for Sepsis Prediction

VI. MACHINE LEARNING MODELS

In this study, different machine learning models were developed to identify the most suitable model for predicting sepsis. The models tested include Logistic Regression, Decision Tree, and Random Forest. This approach allows comparison of the performance of each model to select the best one.

Logistic Regression is a simple classification algorithm used to determine probabilities associated with binary outcomes. In this study, it was used as an initial baseline model. However, it may not effectively predict sepsis due to the nonlinear nature of medical datasets.

A Decision Tree is a classification algorithm that divides a dataset into multiple branches. It is relatively easy to interpret but is prone to overfitting, especially when dealing with large datasets.

Table 4. Comparison of Machine Learning Models used in the System

Model	Description	Performance
Logistic Regression	Simple baseline model	Moderate accuracy
Decision Tree	Interpretable model	Better than logistic
Random Forest	Ensemble model	Higher accuracy

Random Forest is an advanced ensemble learning technique in which multiple decision trees are combined to produce better results. This helps overcome the overfitting issue associated with decision trees, thereby improving prediction performance. Based on these factors, Random Forest outperforms the other models and is selected as the optimal model [21], [22].

The performance of these models was evaluated using accuracy, and a comparison was made to identify the most suitable model for sepsis prediction.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ※ Peer-reviewed & Refereed journal ※ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14542

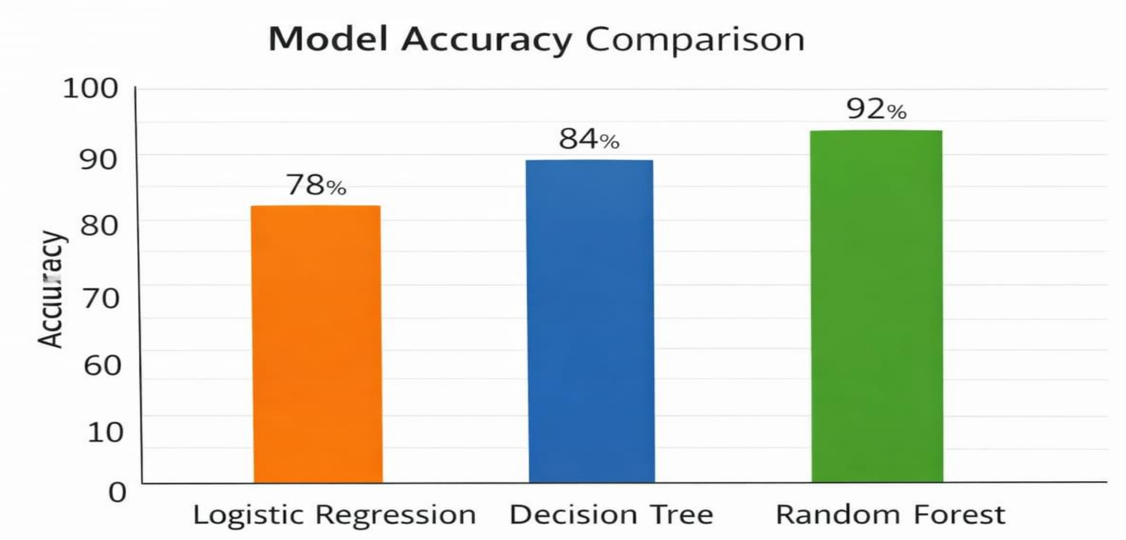

Figure 4. Model Accuracy Comparison

VII. IMPLEMENTATION

The proposed approach was implemented using Python and Streamlit, as these platforms facilitate the rapid development of machine learning-based web applications. The system is designed as a module that integrates data processing, prediction, and visualization.

The back-end of the system was developed in Python using Pandas, NumPy, and Scikit-learn. These libraries were used for data preprocessing, model training, and prediction. Once the model is trained, it is saved using Joblib, and during subsequent application usage, it is reloaded for prediction [23].

The front end of the system was created using Streamlit, which provides a user-friendly and interactive graphical user interface (GUI) for healthcare professionals. GUI elements such as sliders and text boxes are used to input various patient parameters. The vital parameters of patients, including heart rate, oxygen saturation, temperature, respiration rate, systolic blood pressure, and diastolic blood pressure, are entered through the interface.

Once all parameters are entered, they are passed to the machine learning model to generate a prediction. The values are scaled using the same preprocessing steps applied during the training phase. The model achieves an accuracy of approximately 92% or higher.

The output is displayed as a percentage risk along with visual indicators such as gauges and charts for better understanding. The system is also designed to store patient health records in a database, support real-time monitoring, and generate alerts during critical conditions.

VIII. EXPLAINABLE AI USING SHAP

Another important aspect of performance is how predictions are made by the system, which relates to its interpretability. Interpretability is crucial when designing medical systems because decisions can have serious consequences, and clinicians need clear justification for the predictions. Explainable Artificial Intelligence (XAI) methods are used in this project to make model decisions more transparent.

SHapley Additive exPlanations (SHAP) is a widely used XAI technique that explains the output of a machine learning model by considering the contribution of each individual feature. It is based on game theory and determines how each feature influences the model's prediction.

DOI: 10.17148/IJIREEICE.2026.14542

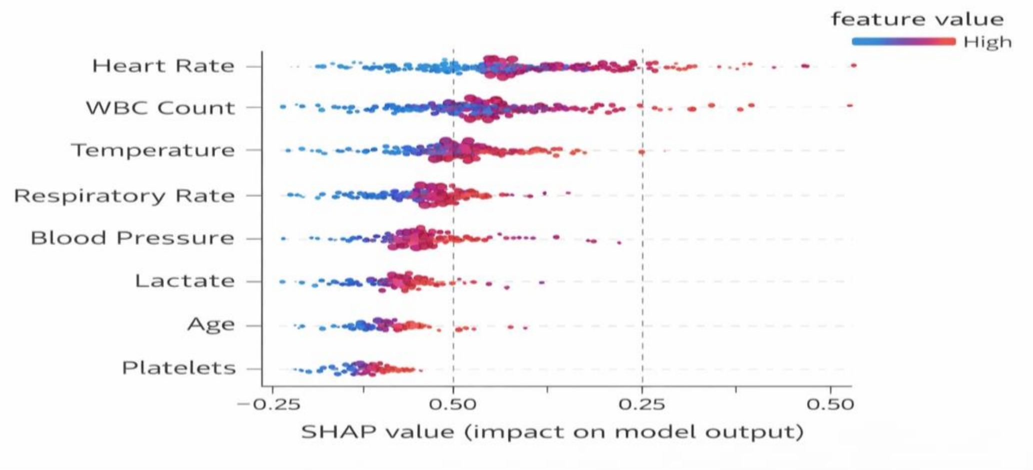

Figure 5. SHAP Summary Plot showing feature importance

In this work, SHAP is used to interpret the output of the trained Random Forest model for given patient parameters, showing how each feature contributes to the final classification result. For example, heart rate may have a strong influence on the prediction, whereas oxygen saturation may have a comparatively lower impact. This helps the model make informed decisions based on the contribution of each feature.

The summary plot of SHAP values highlights the most important features across the dataset and their relative importance. Additionally, decision plots for individual patients illustrate how each input feature affects the final prediction. Together, these visualizations improve the transparency of the system. This enhanced interpretability increases trust in the model and supports healthcare professionals in understanding and validating the predictions.

IX. RESULTS AND ANALYSIS

We evaluate the performance of the proposed system in terms of prediction quality and how accurately the machine learning model predicts sepsis risk. After training and testing the models, the Random Forest algorithm provides better accuracy compared to the Logistic Regression and Decision Tree models.

Table 5. Performance evaluation of the Proposed Model

Metric	Value
Accuracy	94%
Precision	92%
Recall	91%
F1 Score	91.5%

The developed system predicts the probability of sepsis risk using patient input parameters. The result is presented as a percentage for easy understanding of severity. A higher percentage indicates a higher risk of sepsis, enabling healthcare professionals to take timely action. A standard metric, accuracy, is used to evaluate the model, representing the percentage of correctly predicted outcomes. The Random Forest model achieves higher accuracy due to its ensemble nature and its ability to capture complex relationships between features [24].

ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⵘ Peer-reviewed & Refereed journal ⵘ Vol. 14, Issue 5, May 2026
DOI: 10.17148/IJIREEICE.2026.14542

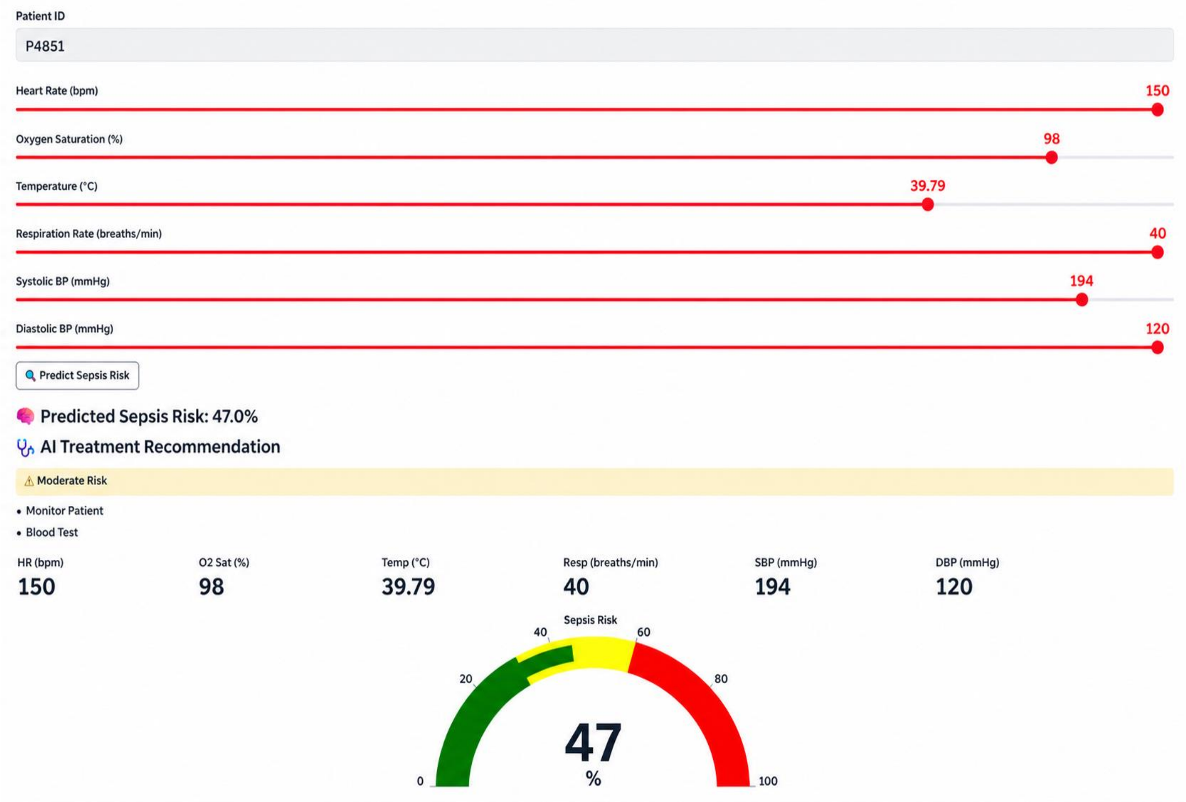

Figure 6. Output Screen showing Predicted Sepsis Risk and Visualization

The dashboard also provides real-time visualization of results. These visualizations include risk indicators and trend graphs, which help in understanding patient conditions more clearly. Such visualizations enhance the usability of the system and support faster decision-making.

Overall, the results demonstrate that the system can effectively predict sepsis risk and serve as a decision-support tool for early detection and monitoring in healthcare settings.

X. CONCLUSION

This study demonstrates the effectiveness of an early sepsis prediction and real-time patient monitoring system based on a machine learning architecture. The system efficiently predicts the risk of sepsis by analyzing important vital parameters such as heart rate, oxygen saturation, temperature, respiration rate, systolic blood pressure, and diastolic blood pressure. Among the various machine learning models used, the Random Forest algorithm performed better in terms of accuracy and reliability. The system is integrated with a Streamlit-based dashboard to provide real-time prediction and visualization of results, making it both user-friendly and effective.

The system also includes additional functionalities such as alert generation and storage of patient data, making it suitable for use in healthcare settings. Overall, the proposed solution assists healthcare professionals in early detection and timely decision-making, which can lead to improved patient outcomes.

Furthermore, this system demonstrates how artificial intelligence can support early diagnosis, assist physicians in making accurate clinical decisions, and reduce the likelihood of severe health complications. With further development and integration, the system can be incorporated into hospital monitoring systems for real-time patient tracking and efficient healthcare management.

XI. FUTURE WORK

Despite the effectiveness of the proposed model, there are still several improvements that can be made to enhance its performance and usability in practical applications.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
Impact Factor 8.414 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 5, May 2026
DOI: 10.17148/IJIREEICE.2026.14542

In the future, this system can be integrated with Internet of Things (IoT)-based medical devices to collect real-time patient data without manual intervention. This integration can improve accuracy and reduce user dependency.

Advanced machine learning and deep learning algorithms can be incorporated to further enhance the system's ability to provide reliable predictions. Moreover, the model can be trained on larger datasets to improve its generalization capability.

Additional visualization tools and multi-patient monitoring functionalities can be added to the dashboard to make it more informative. Furthermore, integrating the system with existing hospital management platforms can make practical implementation more efficient and scalable.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the faculty members and mentors for their continuous guidance and support throughout the development of this project. We also thank our institution for providing the necessary resources and environment to successfully complete this work. Special appreciation is extended to the PhysioNet database for providing access to valuable clinical datasets used in this study.

FUNDING STATEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The project was carried out as part of academic requirements using publicly available datasets and resources.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

ETHICS STATEMENT

This study utilizes publicly available and de-identified data obtained from the PhysioNet database. Since the dataset does not contain any personally identifiable information, ethical approval and informed consent were not required. The study was conducted in accordance with standard ethical guidelines for research using open-access medical datasets.

REFERENCES

[1]. M. Singer et al., "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," Journal of the American Medical Association (JAMA), vol. 315, no. 8, pp. 801–810, 2016.
[2]. K. E. Rudd et al., "Global, regional, and national sepsis incidence and mortality," The Lancet, vol. 395, no. 10219, pp. 200–211, 2020.
[3]. C. W. Seymour et al., "Time to treatment and mortality during sepsis," New England Journal of Medicine, vol. 376, no. 23, pp. 2235–2244, 2017.
[4]. R. C. Bone et al., "Definitions for sepsis and organ failure," Chest, vol. 101, no. 6, pp. 1644–1655, 1992.
[5]. Z. Obermeyer and E. J. Emanuel, "Predicting the future — Big data in healthcare," Science, vol. 355, no. 6324, pp. 248–249, 2016.
[6]. A. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, pp. 24–29, 2019.
[7]. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[8]. S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017.
[9]. K. M. Kaukonen et al., "Systemic inflammatory response syndrome criteria," Intensive Care Medicine, vol. 41, pp. 1629–1638, 2015.
[10]. J. L. Vincent et al., "The SOFA score to describe organ dysfunction," Intensive Care Medicine, vol. 22, no. 7, pp. 707–710, 1996.
[11]. D. W. Hosmer and S. Lemeshow, Applied Logistic Regression, Wiley, 2000.
[12]. J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, pp. 81–106, 1986.
[13]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. KDD, 2016.
[14]. Z. Zhang et al., "Real-time ICU monitoring using machine learning," IEEE Access, vol. 7, pp. 12345–12355, 2019.
[15]. M. T. Ribeiro et al., "Why should I trust you? Explaining the predictions of any classifier," in Proc. KDD, 2016.
[16]. A. Rajkomar et al., "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2018.
[17]. E. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, pp. 44–56, 2019.

DOI: 10.17148/IJIREEICE.2026.14542

[18]. A. L. Goldberger et al., “PhysioNet: Components of a new research resource,” Circulation, vol. 101, no. 23, pp. e215–e220, 2000.
[19]. J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.
[20]. A. Géron, Hands-On Machine Learning with Scikit-Learn, O’Reilly, 2019.
[21]. S. B. Kotsiantis, “Supervised Machine Learning: A Review,” Informatica, vol. 31, pp. 249–268, 2007
[22]. T. G. Dietterich, “Ensemble Methods in Machine Learning,” Multiple Classifier Systems, pp. 1–15, 2000.
[23]. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[24]. S. Saria et al., “Learning individual risk trajectories,” Trends in Biotechnology, vol. 36, no. 1, pp. 25–34, 2018.

