**A PROJECT REPORT ON**

**Accurate Prediction** **of Sepsis in ICU Patients**

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE OF

**BACHELOR OF ENGINEERING**

**IN**

**INFORMATION TECHNOLOGY**

**BY**

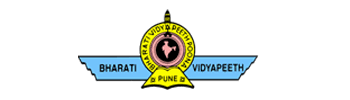
**Sneha Bamane (B190348580)**

**Shweta Maharanawar (B190348541)**

**Devyani Pathrikar (B190348555)**

**UNDER THE GUIDANCE OF**

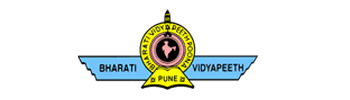
**Prof. M. A. Rane**

****

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**BHARATI VIDYAPEETH'S COLLEGE OF ENGINEERING FOR WOMEN KATRAJ-DHANKAWDI, PUNE - 411043.**

**2023-2024**



# CERTIFICATE

This is to certify that the project report entitled

**Accurate Prediction of Sepsis in ICU Patients**

**Submitted by**

**Sneha Bamane (B190348580)**

**Shweta Maharanawar (B190348541)**

**Devyani Pathrikar (B190348555)**

Is a bonafide work carried out by them under the supervision of Prof. M. A. Rane and it is approved for the partial fulfilment of the requirement of Savitribai Phule Pune University for the award of the Degree of Bachelor of Engineering (Information Technology)

This project report has not been earlier submitted to any other Institute or University for the award of any degree or diploma.

Prof. M. A. Rane Prof. Dr. D. A. Godse

Internal Guide Head of the Department

Department of Information TechnologyDepartment of Information Technology

External Examiner Prof. Dr. P. V. Jadhav

Date:  Principal

BVCOEW

Date:

Place: Pune

## ACKNOWLEDGMENT

To complete any type of project work is teamwork. It involves all the technical/nontechnical expertise from various sources. The contribution from the experts in the form of knows-how and other technical support is of its vital importance. We are indebted to our inspiring guide **Prof. M. A. Rane,** our project coordinator.

**Prof. Dr. K. A. Malgi** and our motivating panel members **Prof. S. A. Hadke** who have extended valuable guidelines, help and constant encouragement through the various difficult stages for the onslaught of the project. We are also very thankful to **Prof. Dr. D. A. Godse** for her valuable guidance.

We have great pleasure in offering big thanks to our honourable Principal **Prof. Dr. P. V. Jadhav.** Last but not least, we would like to thank all the direct and indirect help provided by the staff and our entire class for successful completion of this project.

Sneha Bamane

Shweta Maharanawar

Devyani Pathrikar

# ABSTRACT

The "Accurate Prediction of Sepsis in ICU Patients" initiative is a comprehensive approach to addressing sepsis, a severe and often life-threatening condition frequently encountered in intensive care units (ICUs). This initiative combines an awareness campaign with advanced predictive modeling to improve early detection and intervention. The awareness campaign aims to educate the public and healthcare professionals about sepsis, emphasizing its severity, common symptoms—such as fever, increased heart rate, rapid breathing, confusion, and severe discomfort—and the critical need for prompt medical attention. By empowering individuals to recognize these signs early and seek immediate help, the campaign seeks to improve patient outcomes significantly.

Concurrently, the initiative employs advanced machine learning techniques, particularly the Random Forest algorithm, to develop a predictive model for sepsis. This involves comprehensive data preprocessing to ensure data quality, addressing class imbalances to avoid biased predictions, and incorporating clinical assessment tools like the Sequential Organ Failure Assessment (SOFA) and quick SOFA (qSOFA) criteria. These steps enhance the model's accuracy in identifying patients at high risk of sepsis, enabling timely and targeted interventions that can be life-saving.

A dedicated website supports this dual approach by serving as a central hub for sepsis education and the dissemination of the predictive model. It offers valuable resources, including educational materials and guidelines for recognizing sepsis symptoms, and provides healthcare professionals with access to the predictive model for clinical use.

In summary, the "Accurate Prediction of Sepsis in ICU Patients" initiative aims to raise sepsis awareness, improve predictive capabilities, and support healthcare professionals and the public in combating sepsis, contributing to a healthier and more informed future.

Keywords: Sepsis, Random Forest, Intensive care units (ICUs), Sequential Organ Failure Assessment (SOFA), quick SOFA (qSOFA).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | **CONTENTS** |  |
| Acknowledgement | | |  | III |
| Abstract | | |  | IV |
| List of Abbreviations | | |  | V |
| List of Figures | | |  | VI |
| List of Graphs | | |  | VII |
| List of Tables | | |  | VIII |
| **Sr. No.** | | | **Chapter** | **Page No.** |
| **1.** |  | **Introduction to Project** | | 1-5 |
|  | 1.1 | Introduction to Project | | 1-2 |
|  | 1.2 | Aims / Motivation behind Project | | 2 |
|  | 1.3 | Overview of the Project | | 3 |
|  | 1.4 | Need of the Project | | 3-4 |
|  | 1.5 | Organization of the report | | 4-5 |
| **2.** |  | **Literature Survey** | | 6-8 |
|  | 2.1 | Introduction | | 6 |
|  | 2.2 | Related Work | | 6-8 |
| **3.** |  | **Problem Statement** | | 9-11 |
|  | 3.1 | Features of the Project | | 9-10 |
|  | 3.2 | Scope of the Project | | 10 |
|  | 3.3 | Objectives of the Project | | 10 |
|  | 3.4 | Constraints of the Project | | 11 |
| **4.** |  | **Project Requirements** | | 12-13 |
|  | 4.1 | H/W and S/W Requirements | | 13 |
| **5.** |  | **System Analysis of Proposed Architecture** | | 14-21 |
|  | 5.1 | Flowchart / Proposed System Architecture | | 14-16 |
|  | 5.2 | High Level Design of the Project | | 16-21 |
|  |  | 5.2.2 Data Flow Diagrams up to Level 2 | | 16-18 |
|  |  | 5.2.3 UML Diagrams: Use Case, Class Diagram, Activity Diagram … | | 19-21 |
| **6.** |  | **System Implementation** | | 22-25 |
|  | 6.1 | Algorithm Style | | 22-24 |

6.2 Description of Detailed Methodologies 24-25

###### Test Cases 26

###### Proposed GUI/Working modules/Experimental Results

27-37

###### Conclusion 38

###### Bibliography in IEEE format Appendices

* 1. Plagiarism Report of Paper and Project report from any open-source tool.
  2. Base Paper(s)
  3. Tools used / Hardware Components specifications.
  4. Published Papers and Certificates.

# LIST OF ABBREVIATIONS

1. AAA – Authentication Authorization, Accounting
2. ACK- Acknowledgement
3. ALU- Arithmetic and Logical Unit
4. BASIC - Beginners All-Purpose Symbolic Instruction Code
5. ML - Machine Learning
6. SOFA - Sequential Organ Failure Assessment
7. qSOFA - Quick Sequential Organ Failure Assessment

# LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Figure Name** | **Page No.** |
| **1.1** | **Agile Development Model** | **3** |
| **5.1** | **Proposed System Architecture** | **14** |
| **5.2** | **Level 0 DFD** | **16** |
| **5.3** | **Level 1 DFD** | **17** |
| **5.4** | **Level 2 DFD** | **18** |
| **5.5** | **Use Case Diagram** | **19** |
| **5.6** | **Class Diagram** | **20** |
| **5.7** | **Activity Diagram** | **21** |

**LIST OF TABLES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** |  | **Table Name** | **Page No.** |
| **7.1** | **Test case** |  | **25** |

# CHAPTER 1 INTRODUCTION TO PROJECT

###### Introduction to Project

Sepsis, a life-threatening condition triggered by infections, remains a critical concern within the realm of intensive care units (ICUs). Our project " Accurate Prediction of Sepsis in ICU Patients" tackle this issue. This approach provides two benefits: raising awareness about sepsis and predictive model for detection of sepsis.

At its core, prioritizes education and awareness. Our initiatives target both the general public and healthcare professionals because we understand how crucial knowledge is in combating sepsis. By educating people about the seriousness of sepsis and the critical need for early detection and intervention, we aim to enable individuals to identify its symptoms and seek prompt medical care.

In parallel, the project leverages advanced machine learning techniques, specifically Random Forest, to construct a predictive model for sepsis. The model's rigorous fine-tuning is aimed at achieving accurate identification of sepsis risk in ICU patients. This predictive system represents advancement in healthcare, offering early detection and improved outcome.

Our data-driven approach encompasses comprehensive data preprocessing, addressing class imbalances within the dataset. Additionally, incorporating the Sequential Organ Failure Assessment (SOFA) score and the quick SOFA (qSOFA) criteria improves predictive accuracy by assessing organ failure trajectories and risk factors. Furthermore, the project features a dedicated website that serves as a vital platform for sepsis education and for distributing the predictive model to the medical community.

###### Aims/Motivation behind Project

The driving force behind the Project can be distilled into a few key motivations:

* + 1. **Sepsis Severity**: The project is motivated by the gravity of sepsis as a life-threatening condition, especially within intensive care units (ICUs). Recognizing the severity of sepsis fuels our commitment to addressing this critical issue.
    2. **Early Detection**: Early detection and intervention are pivotal in sepsis management. We are motivated to raise awareness and create a predictive model to enable early sepsis identification, ultimately saving lives.
    3. **Data-Driven Solutions**: We are driven by the belief in data-driven solutions. Leveraging advanced machine learning techniques, we seek to use data to enhance decision-making in healthcare.
    4. **Patient-Centered Care**: Our core motivation is patient-centered care. We are dedicated to improving patient outcomes, reducing sepsis- related burdens, and contributing to a healthier future.
    5. **Interdisciplinary Approach**: Recognizing the value of interdisciplinary methods, we are motivated to address sepsis comprehensively by combining education, technology, and clinical insights.
    6. **Raising Awareness**: The first step in combating sepsis is awareness. We are motivated to spread knowledge about sepsis, ensuring both the public and healthcare professionals are equipped to recognize and respond to this condition.
    7. **Reducing Tragedies**: Our ultimate motivation is to reduce sepsis- related tragedies. The knowledge that sepsis is a leading cause of mortality in ICUs drives us to integrate clinical insights, including qSOFA, into our predictive model to make a tangible difference.

###### Overview of the Project

Our Project is executed following an Agile software development model. This model ensures flexibility, adaptability, and a continuous improvement

approach. The project unfolds through iterative cycles, allowing for regular feedback and enhancements. Each cycle includes phases of requirement gathering, system design, implementation, testing, and deployment. Data preprocessing and modeling are central components, with Random Forest regression driving predictive capabilities. Integration of the Sequential Organ Failure Assessment (SOFA) score, including qSOFA criteria, enhances accuracy. Additionally, the project maintains a dedicated website for sepsis education and predictive model dissemination. This approach enables the project to remain responsive to evolving needs, deliver incremental value, and align with its overarching goals effectively.

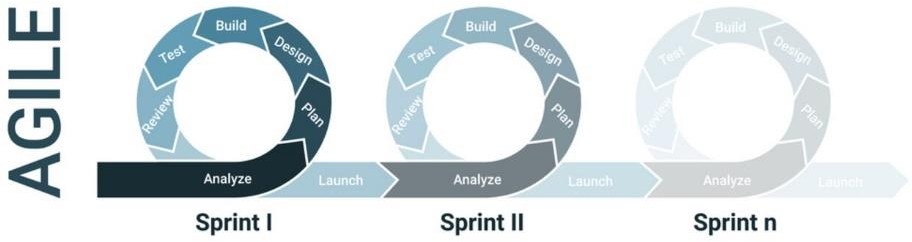


Fig.1.1 Agile software development model

###### Need of the Project

Our Project addresses a pressing need within the healthcare landscape. Sepsis, a life-threatening condition, demands swift identification and intervention, particularly in intensive care units (ICUs) where patients are most vulnerable. The significance of this study lies in its potential to revolutionize patient care by bridging critical gaps. By fostering sepsis awareness among both the public and healthcare professionals, the project empowers individuals to recognize symptoms early, ultimately saving lives. Furthermore, the development of an advanced predictive model for sepsis, integrated with the Sequential Organ Failure Assessment (SOFA) score and qSOFA criteria, ensures rapid and accurate risk assessment. This innovation enhances patient outcomes and contributes to the reduction of sepsis-related fatalities in ICUs. The project's profound impact extends beyond the boundaries of sepsis, serving as a model for the successful integration of education, data-driven technology, and clinical insights in healthcare. Ultimately, the significance of this study lies in its

potential to create a healthier, more informed future, where sepsis-related tragedies are minimized, and the standards of patient care in ICUs are elevated.

###### Organization of the Project Report

* + 1. **Introduction to Project:**
       - Provide an overview of the project, introducing the significance of early sepsis detection in the ICU setting and the purpose of the Project.

###### Aims / Motivation behind Project:

* + - * Explain the motivation for your project, which is to improve the accuracy of sepsis prediction using machine learning and enhance patient outcomes in ICUs.

###### Overview of the Project:

* + - * Outline the main components of the project, including the utilization of SOFA and qSOFA scores, machine learning algorithms, and the goal of achieving early sepsis detection.

###### Need of the Project:

* + - * Explore the reasons why accurate sepsis prediction in ICU patients is essential. Discuss the challenges and gaps in current sepsis detection methods that your project aims to address.

###### Literature Survey (minimum five papers):

* + - * Provide a literature review section that covers relevant academic and research papers. Discuss the background, related work, and the problem statement in the context of sepsis prediction.

###### Features of the Project:

* + - * Describe the core features of the project, focusing on the application of SOFA and qSOFA scores, machine learning models, and their integration into clinical practice.

###### Scope of the Project:

* + - * Define the boundaries of your project, clarifying what aspects of sepsis detection in ICUs are covered. This sets expectations for the reader.

###### Objectives of the Project:

* + - * List the specific goals and objectives of your project, highlighting what you intend to achieve with regard to sepsis prediction.

###### Constraints of the Project:

* + - * Address any limitations or constraints that may have influenced the development of the project, such as data availability or model performance.

###### Project Requirements:

* + - * Outline the hardware and software requirements necessary to implement and run project. This section explains what's needed for the platform to function effectively.

###### System Analysis of Proposed Architecture:

* + - * Delve into the technical aspects of your project, including system architecture, data flow diagrams, and any relevant UML diagrams. This provides a technical understanding of how the platform operates.

###### System Implementation:

* + - * Discuss the technical implementation of your sepsis prediction system, including the use of machine learning algorithms, data preprocessing, and the integration of SOFA and qSOFA scores.

###### Project Plan:

* + - * Present the project's timeline, milestones, and development plans. This section outlines how the project will be executed.

###### Conclusion:

* + - * Summarize the project's key takeaways, achievements, and its potential impact on improving sepsis detection in ICU patients. Reflect on the project's journey and outcomes.

# CHAPTER 2 LITERATURE SURVEY

###### Introduction

This literature survey provides an overview of relevant research in the field of sepsis prediction, emphasizing the importance of early diagnosis and the application of machine learning techniques. By reviewing recent studies, we aim to establish a contextual framework for our project. These studies highlight the significance of predictive models, such as Random Forest, Logistic Regression, and Gradient Boosting, in predicting sepsis and in-hospital mortality. Through this literature survey, we aim to gain insights from existing work and inform the direction of our research.

* 1. **Related Work**

|  |  |  |  |
| --- | --- | --- | --- |
| **Title** | **Author** | **Publication** | **Remark** |
| Early Prediction of Sepsis using Machine Learning | Anurag Shankar, Mufaddal Diwan, Snigdha Singh, Husain Nahrpurawala and Tanusri Bhowmick | IEEE, 2021 | The paper introduces the importance of the SOFA score and the challenges in timely diagnosis, serving as a basis for our project. |
| A Comprehensive Machine Learning Based Pipeline for an Accurate Early Prediction of Sepsis in ICU | B. C. SRIMEDHA, RASHMI NAVEEN RAJ, AND VEENA MAYYA | IEEE  Access, 2022 | This research by investigating four prediction algorithms, including Random Forest, Logistic Regression, Gradient Boosting, and Decision Tree, and examining the impact of various imputation techniques, |

|  |  |  |  |
| --- | --- | --- | --- |
| Using machine learning methods to predict in- hospital mortality of sepsis patients in the ICU | Guilan Kong, Ke Lin and Yonghua Hu | 2020,  Published by BMC. | This study focuses on leveraging machine learning techniques to predict the in-hospital mortality of sepsis patients in the ICU. Machine learning models, including the least absolute shrinkage and selection operator (LASSO), random forest (RF), gradient boosting machine (GBM), and traditional logistic regression (LR), were developed for prediction. |
| A Machine Learning Model for Early Prediction and Detection of Sepsis in Intensive Care Unit Patients | Yash Veer Singh, Pushpendra Singh, Shadab Khan, and Ram Sewak Singh | 2022,  Published by Hindawi. | This paper introduces a machine learning model for early sepsis prediction in ICU patients, leveraging data from clinical laboratory values and vital signs. Various models, including SVM, RF, NB, LR, and XGBoost, are examined and compared, with the proposed ensemble method showing the most promising results in terms of classification performance and prognosis improvement. |
| Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models | Longxiang Su, Zheng Xu, Fengxiang Chang, Yingying Ma, Shengjun Liu, Huizhen Jiang, Hao Wang, Dongkai Li, Huan Chen, Xiang Zhou, Na Hong, Weiguo Zhu, and Yun Long | 28 June  2021,  Published by Frontiers in Medicine | This study centers around harnessing machine learning techniques to anticipate in- hospital mortality among sepsis patients in the ICU. Several machine learning models were constructed for this purpose, encompassing the least absolute shrinkage and selection operator (LASSO), random forest (RF), gradient boosting machine |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | (GBM), and the conventional logistic regression (LR). |

# CHAPTER 3 PROBLEM STATEMENT

To develop a predictive model for early detection of sepsis in ICU patients and to create a website for generating a website for sepsis awareness containing information about sepsis.

#### Features of the Project:

* + - **Sepsis Awareness Campaign**: The project includes a dedicated website and educational resources to raise awareness about sepsis, its symptoms, and the importance of early detection.
    - **Educational Content**: Informative materials simplify complex medical information, making it accessible to the general public and healthcare professionals, empowering them to recognize sepsis symptoms.
    - **Advanced Predictive Modeling**: The project leverages Random Forest regression for the development of a predictive model that accurately identifies sepsis risk in ICU patients.
    - **Integration of qSOFA Criteria**: The predictive model integrates the quick Sequential Organ Failure Assessment (qSOFA) criteria, enabling rapid risk assessment and intervention.
    - **Data Preprocessing**: Comprehensive data preprocessing techniques are applied to address class imbalances in the dataset, enhancing the model's predictive accuracy.
    - **Interdisciplinary Approach**: The project combines education, data-driven technology, and clinical insights, ensuring a holistic approach to sepsis management.
    - **Digital Platform**: The project maintains a dedicated website as a central platform for sepsis education and the dissemination of the predictive model to the medical community.
    - **Continuous Improvement**: An agile development model allows for flexibility and ongoing enhancements to meet evolving needs effectively.
    - **Data-Driven Insights**: The project ensures that data-driven insights are used to inform clinical decisions and enhance early intervention strategies.

#### Scope of the Project:

The scope of our project encompasses creating awareness, implementing advanced predictive modeling for sepsis detection, and fostering a holistic approach to sepsis management. It aims to reduce sepsis-related mortality rates in ICUs, improve patient outcomes, and serve as a model for integrating education and technology in healthcare practices.

#### Objectives of the Project:

* + - To educate the public and healthcare professionals about sepsis, its severity, and the importance of early recognition.
    - To develop and implement an advanced predictive model for sepsis to enable early detection and intervention in intensive care units (ICUs).
    - To improve patient outcomes by minimizing sepsis-related fatalities through early intervention and rapid risk assessment.
    - To utilize data-driven insights to inform clinical decisions and enhance the quality of patient care in ICUs.
    - To combine education, technology, and clinical insights to create a holistic approach to sepsis management.
    - To contribute to a reduction in the burden of sepsis by fostering proactive healthcare practices.
    - To serve as a model for successfully integrating education and technology into healthcare practices, potentially impacting other critical medical conditions.

#### Constraints of the Project:

* + - **Data Availability**: The project relies on the availability of comprehensive and reliable healthcare data for the development of the predictive model, which can sometimes be limited.
    - **Technological Infrastructure**: Access to suitable technological infrastructure and resources is crucial for implementing the predictive model and maintaining the project's digital platform.
    - **Clinical Adoption**: The successful adoption of the predictive model within clinical settings may face resistance or challenges, requiring careful integration strategies.
    - **Data Privacy and Ethics**: Adhering to data privacy and ethical considerations in healthcare data usage is a paramount constraint that demands rigorous compliance.

# CHAPTER 4 PROJECT REQUIREMENTS

Project Requirements are as follows:

* + - **Data Set**: Access to a comprehensive healthcare data set, sourced from Kaggle, is crucial for the development and training of the predictive model.
    - **Algorithm**: Implementing the Random Forest predictive model, ensuring its accuracy.
    - **HTML and CSS**: Develop a user-friendly web interface using HTML and CSS to provide educational resources, awareness materials, and model access to users.
    - **Flask Framework**: Utilize the Flask web framework to build the project's digital platform, integrating the educational content, predictive model, and awareness campaigns.
    - **Database**: Set up a database to store and manage healthcare data required for the predictive model, ensuring data security, privacy, and efficient retrieval.
    - **Technology Infrastructure**: Ensure access to the necessary technological infrastructure, including server resources, for hosting the project's website and predictive model.
    - **Budget and Funding**: Secure adequate funding to sustain the project, covering the costs associated with awareness campaigns, server hosting, and development efforts.
    - **Data Privacy and Ethics**: Adhere to stringent data privacy and ethical standards when handling sensitive healthcare data within the project.
    - **Flexibility**: Maintain a flexible project approach to accommodate the evolving landscape of medical practices and technological advancements.

By fulfilling these requirements, the project can effectively bridge the gap in sepsis awareness and early detection while contributing to improved patient care in ICUs.

###### Hardware and Software Requirements

· **Hardware**

* + 1. Computer(s)
    2. System: Intel I3 processor or above
    3. RAM: 4GB or above

###### Software

1. Operating System: Windows 10 or any open-source OS
2. Programming Languages: Python for machine learning, data development.
3. Development Tools: Visual Studio Code, Anaconda

###### Other equipment

1. Internet Connection
2. Data Sources

# CHAPTER 5

**SYSTEM ANALYSIS OF PROPOSED ARCHITECTURE**

###### Proposed System Architecture

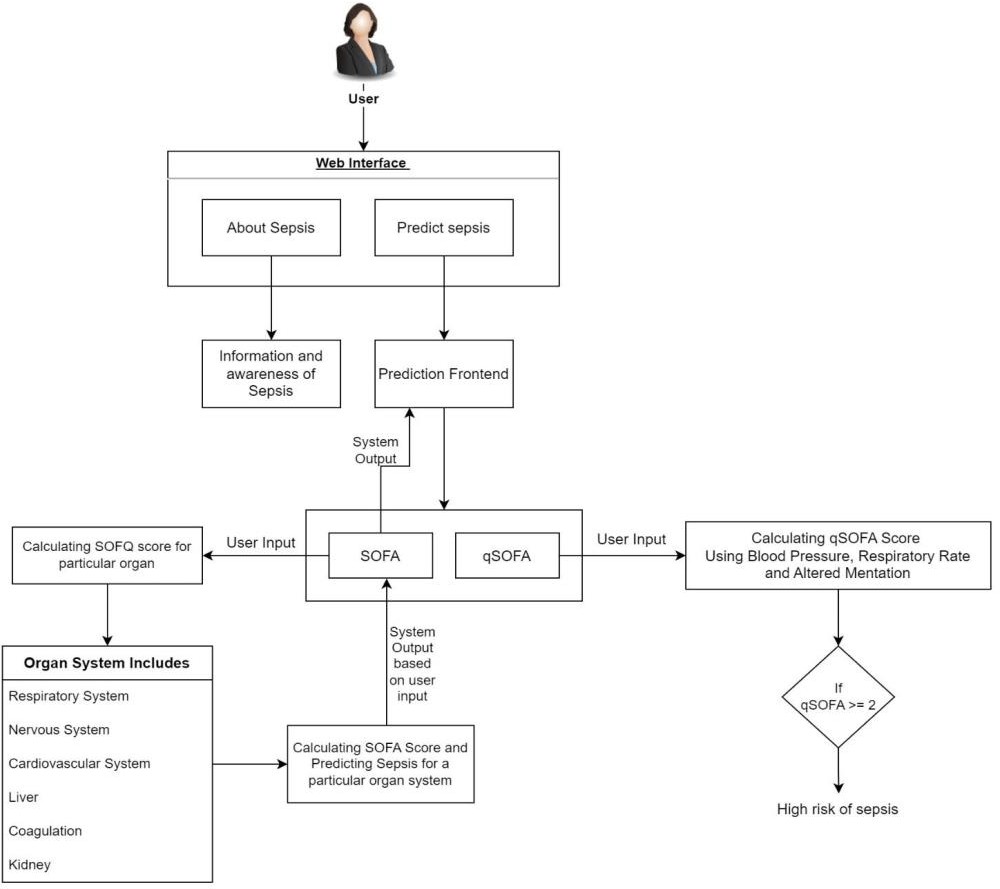


Fig 5.1 Proposed System Architecture

The system architecture for the **Accurate Prediction of Sepsis in ICU Patients** is designed to integrate technology, education, and clinical insights to address the issue of sepsis in intensive care units (ICUs).

###### Input Data:

* + The architecture comprises a set of vital signs and clinical parameters from ICU patients as input data. Specifically, we focus on:
    - Low (Diastolic) Blood Pressure
    - High (Diastolic) Blood Pressure
    - O2 Saturation
    - Altered Mentation (Glasgow Coma Scale)
    - Respiratory Rate (Breaths per minute)
    - Temperature
    - Heart Rate (Beats per minute)
    - Respiratory Rate (PaO2/FiO2)
    - Liver (Bilirubin)
    - Kidney Function (Creatinine)

###### SOFA Score Calculation:

* + The SOFA score is a comprehensive assessment that evaluates the patient's condition across various organ systems. This includes:
    - Respiratory System
    - Liver
    - Kidney
  + Machine learning algorithms are employed to calculate the SOFA score by considering the values of these organ systems.

###### qSOFA Score Calculation:

* + In contrast to SOFA, the qSOFA score is a quicker assessment.
    - Low Blood Pressure (SBP <= 100mmHg)
    - Respiratory Rate (>= 22 breaths /min)
    - Altered Mentation (GCS <= 14)
  + We use machine learning techniques to calculate the qSOFA score, which allows quick predictions of sepsis risk.

###### Prediction Outcome:

* + Both the SOFA and qSOFA scores play a crucial role in our sepsis prediction. These scores are used as features to train machine learning models. This integrates the scores with other clinical parameters to enhance accuracy.

###### Machine Learning Algorithms:

* + While we use machine learning algorithms to calculate the scores, we also employ additional machine learning models for sepsis prediction. These models are fine-tuned for accuracy and will be chosen based on their performance. Future sections of the project report will provide detailed information on the specific machine learning algorithms used.

Our architecture allows us to leverage clinical data, vital signs, and organ system assessments to make informed predictions about a patient's risk of developing sepsis.

By using SOFA and qSOFA scores as integral components, we ensure a comprehensive and timely approach to sepsis prediction in ICU patients.

###### High Level Design of the Project

* + 1. **Data Flow Diagrams**

###### Level 0 DFD

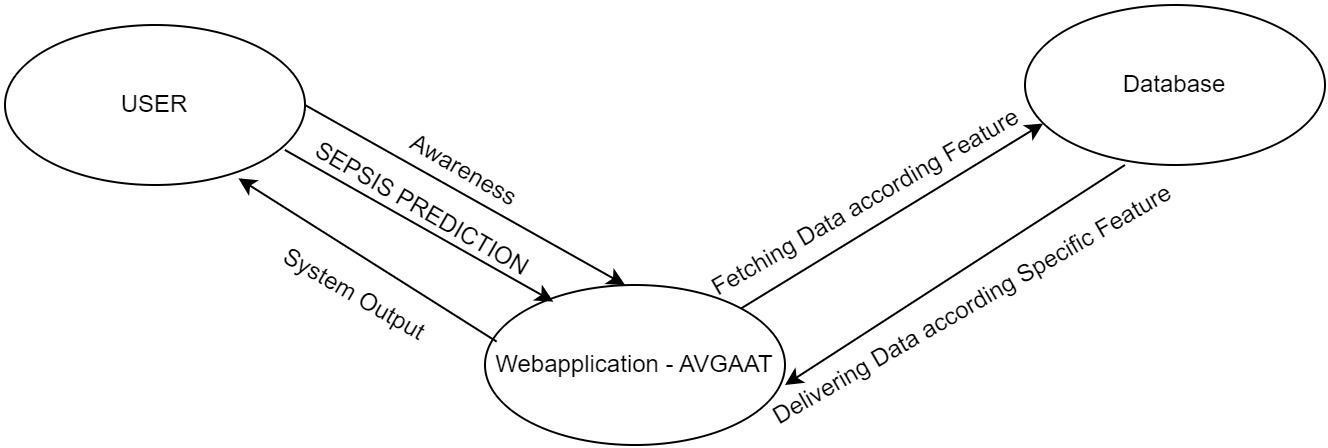


Fig 5.2 Level 0 DFD

###### Level 1 DFD

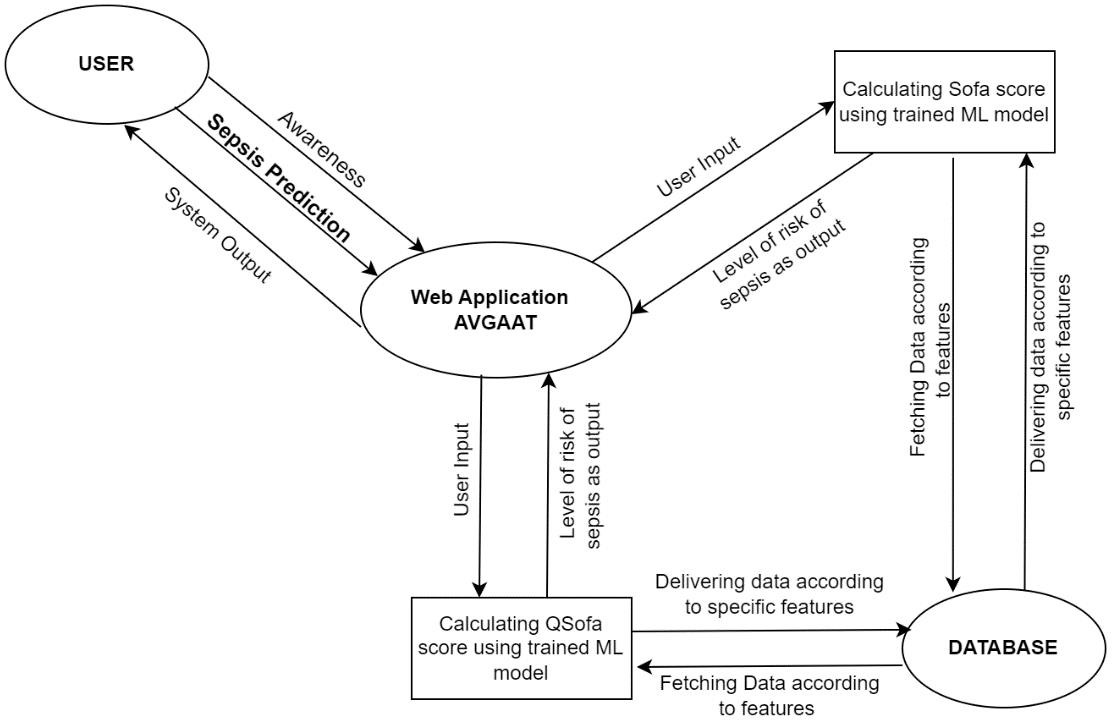


Fig 5.3 Level 1 DFD

###### Level 2 DFD

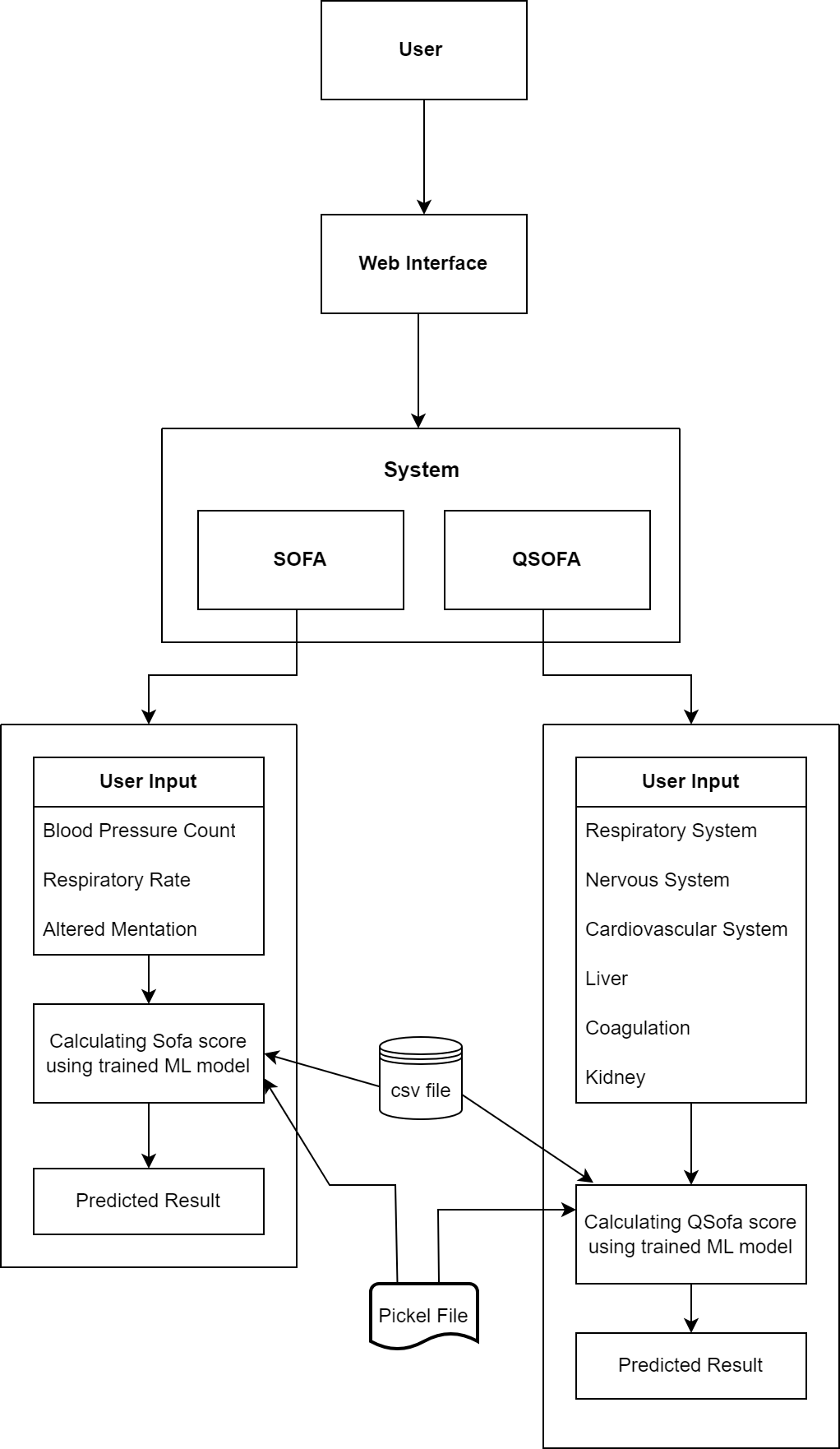


Fig 5.4 Level 2 DFD

###### UML Diagrams

**Use Case Diagram**

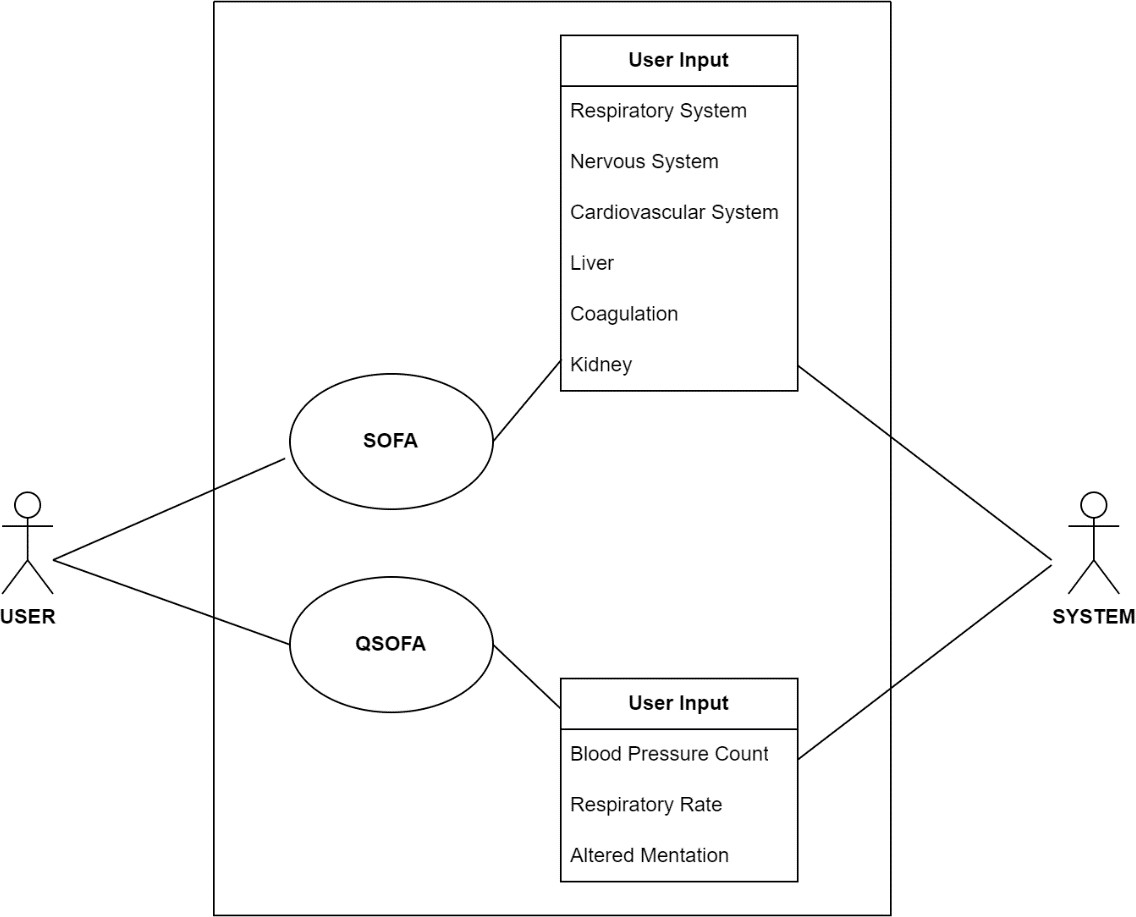


Fig 5.5 Use Case Diagram

###### Class Diagram

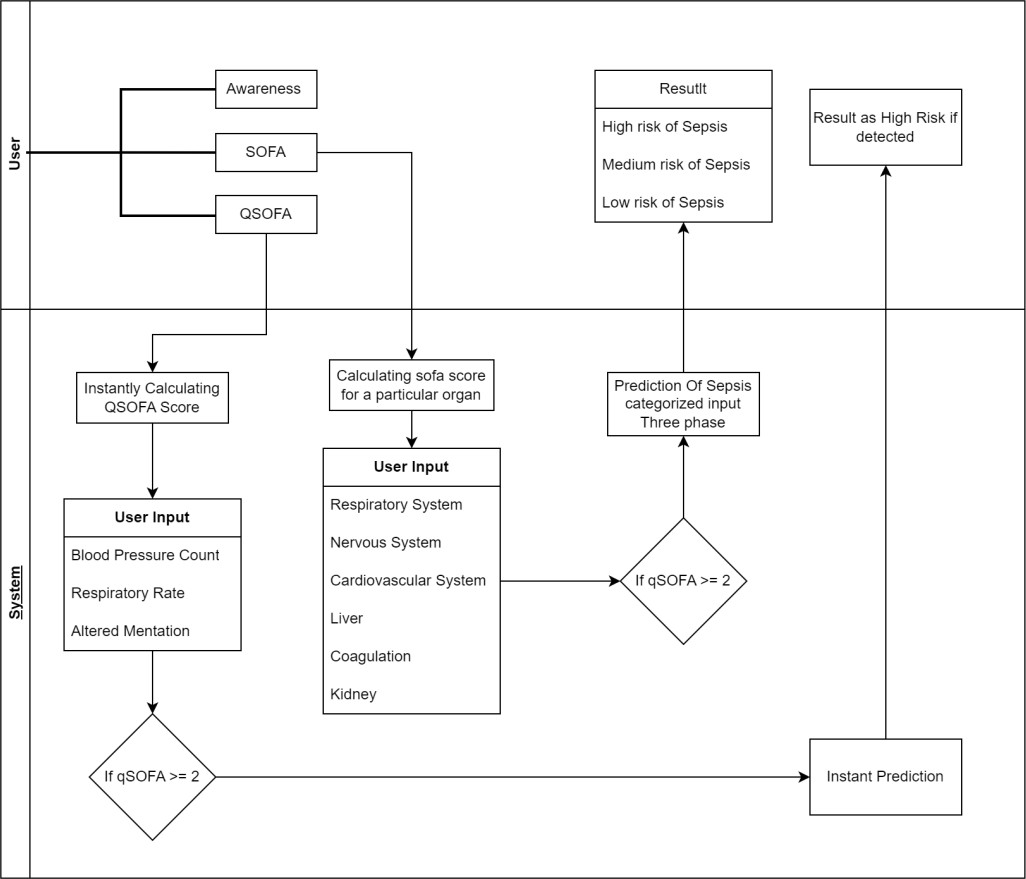


Fig 5.6 Class Diagram

###### Activity Diagram

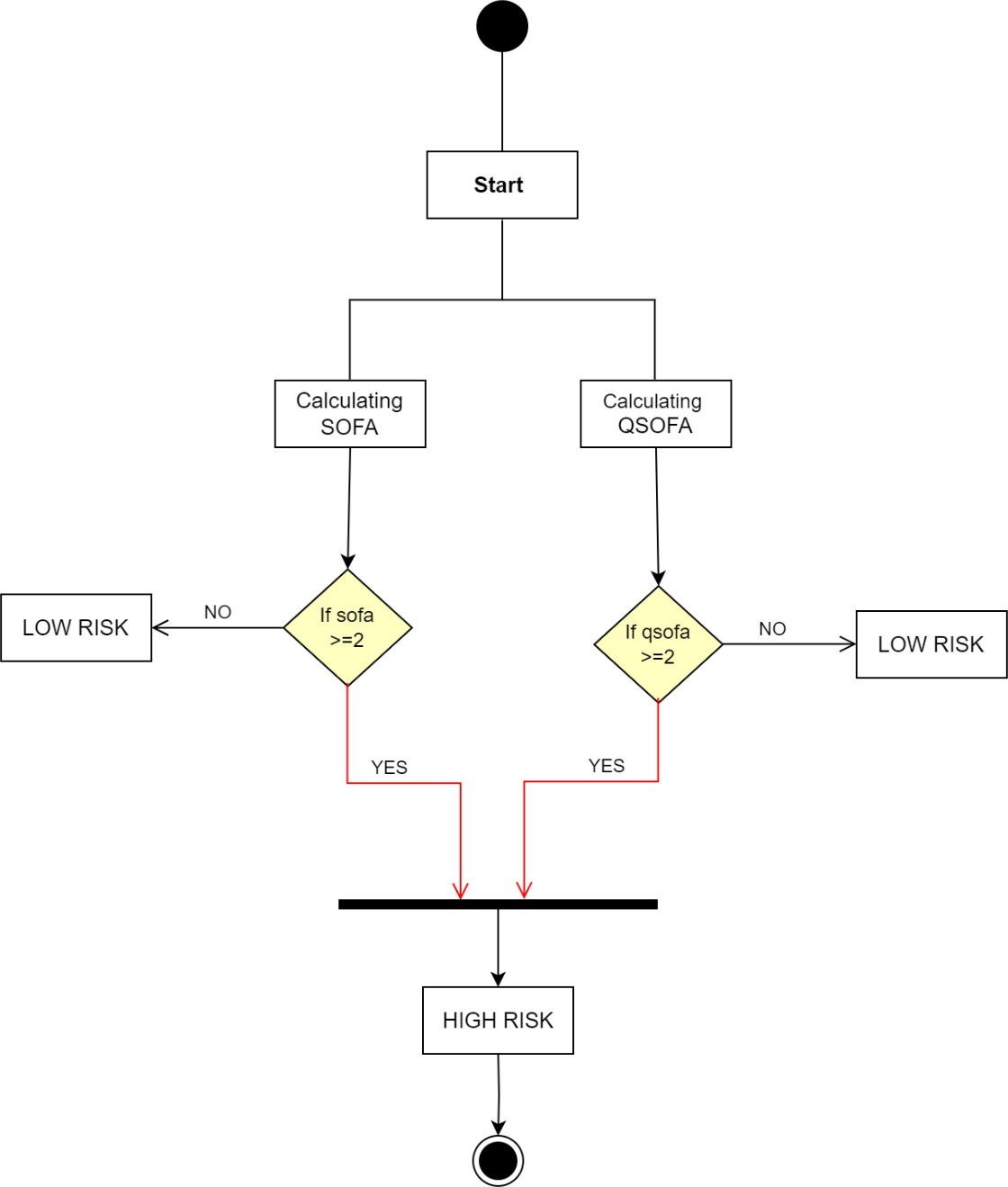


Fig 5.7 Activity Diagram

# CHAPTER 6 SYSTEM IMPLEMENTATION

###### Algorithm Style

* + 1. **Random Forest**

###### Bootstrap Aggregating (Bagging)

Each decision tree in a Random Forest is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. For example, if our training set has 1000 rows, each decision tree gets a sample of 1000 rows drawn randomly from the training set, allowing the same row to appear multiple times in each sample.

###### Feature Randomness

When building each tree, each time a split in a tree is considered, a random subset of the features is chosen as split candidates from the full set of features. This is called the random subspace method. The purpose is to decrease the correlation between the trees in the forest. Typically, for a classification problem, the square root of the total number of features is used, whereas for algorithm, one-third of the total number of features is chosen.

###### How It Works

Each tree gives a prediction in the case of regression, the average of all the trees’ predictions is the final prediction. For classification, the prediction is the class with the majority votes across all trees.

###### Internal working of algorithm:

**Step 1: Bootstrapping the Data**

For each tree in the forest, algorithm starts by performing bootstrapping, which is a resampling technique. In this process, random samples of the training dataset are drawn with replacement. This means some observations may be repeated in each bootstrap sample, while others may be left out. Each bootstrap sample is used to train a separate decision tree, which helps ensure that the trees in the forest are de-correlated.

###### Step 2: Building Decision Trees

Random Forests build decision trees slightly differently from the usual method:

* 1. **Random Feature Selection**: When creating a split at a node in a decision tree, Random Forest randomly selects a subset of the features rather than considering every possible feature. The size of this subset can be controlled by parameters such as **max\_features** in scikit-learn. For classification tasks, this is typically the square root of the total number of features, and for regression, it’s about one-third. This randomness helps to make the model more robust and prevents overfitting.
  2. **Node Splitting**: Each tree grows during the training process by splitting nodes based on the selected features. The splits are chosen to best separate the classes in classification or minimize the variance in regression. This is typically done using measures such as Gini impurity or entropy for classification and mean squared error for regression.
  3. **Tree Depth**: Trees in algorithm are generally grown to their full depth, but parameters such as **max\_depth** can limit this. Full trees can mean that each tree can potentially overfit the data it was trained on, but when aggregated, the forest as a whole does not overfit as much.

###### Step 3: Aggregation of Predictions

After all trees are trained:

* **Classification**: Each tree in the forest outputs a class prediction. The final prediction of the Random Forest is the class that receives the majority vote among all the trees.
* **Regression**: Each tree predicts a continuous value. The final prediction of the Random Forest is typically the average of these predictions.

###### Step 4: Handling Overfitting and Variance

It handles overfitting through two mechanisms:

1. **Bootstrap Aggregating (Bagging)**: Since each tree is trained on a different sample of data, the variance of the model is reduced. Even if some trees overfit their samples, the averaging makes the overall model less prone to overfitting.
2. **Feature Randomness**: By forcing each split to consider only a subset of the features, the trees are made less correlated with each other. Less correlation among trees in the model means a lower chance that they will all repeat the same errors, thus reducing the model variance further.

In our project, Random Forest can do various task like:

* **Predictive Modelling:** It is a supervised learning and bagging technique that uses an ensemble learning method for prediction in machine learning. The trees here run in parallel, meaning there is no interaction between these trees while building the trees. Random Forest can build a predictive model aimed at determining the likelihood of a patient developing sepsis during their ICU stay, based on diverse patient-related factors and clinical data.
* **Feature Selection:** It can help identify which patient variables or features are most relevant for predicting sepsis. This involves selecting and weighting various clinical and laboratory measurements, vital signs, and other patient data that are considered potential risk factors for sepsis.
* **Risk Stratification:** Random forest regression can help in stratifying patients into different risk categories based on their calculated sepsis risk scores. This can assist healthcare providers in prioritizing care and interventions for patients at higher risk of sepsis.

###### Description of Detailed Methodologies Data Acquisition and Preprocessing:

It begins by collecting healthcare data from a dataset sourced from Kaggle. This data is then preprocessed to handle class imbalances and ensure its suitability for predictive modeling.

###### Predictive Modeling:

We've fine-tuned algorithm to improve early sepsis detection accuracy. The predictive model developed using Random Forest has yielded results in forecasting the likelihood of patients developing sepsis during their ICU stay. With an impressive accuracy rate of 85%, this model represents a significant advancement in sepsis prediction, providing

healthcare professionals with valuable insights for proactive intervention and patient management.

###### Web Interface:

It maintains a user-friendly web interface built with HTML and CSS, providing access to educational content and the predictive model. Users, including the general public and healthcare professionals, can access information on sepsis awareness, symptoms, and risk assessment.

###### Flask Web Framework:

It serves as the backbone of the digital platform. It handles user requests, routes them to the components, and ensures better communication between the user interface and the server.

###### Interdisciplinary Collaboration:

Effective collaboration between healthcare professionals, data scientists, web developers, and educators is pivotal. It ensures that the project combines expertise in medical practices, data analysis, web development, and educational content.

###### Security and Privacy Compliance:

The system adheres to stringent security and data privacy standards, safeguarding sensitive healthcare information and ensuring ethical data usage.

# CHAPTER 7 TEST CASES

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Functional Test Cases | Actual Output | Expected Output | Test Status |
| 1 | Verify if the User is able to see the complete UI of the system. | The UI is visible. | The UI is visible. | Pass |
| 2 | Verify if sepsis s detected with low risk. | Sepsis detected with low risk. | Sepsis detected with low risk. | Pass |
| 3 | Verify if sepsis s detected with medium risk. | Sepsis detected with medium risk. | Sepsis detected with medium risk. | Pass |
| 4 | Verify if sepsis s detected with high risk. | Sepsis detected with high risk. | Sepsis detected with high risk. | Pass |
| 5 | Verify Admin login working properly. | Works properly. | Works properly. | Pass |
| 6 | Verify user data is saved in database. | Data saved successfully. | Data saved successfully. | Pass |
| 7 | Verify if the training model is giving good accuracy or not | Showed Accuracy of 85% | Should show accuracy of minimum 80% | Pass |

Table 7.1. Test case

# CHAPTER 8

**PROPOSED GUI/WORKING MODULES/EXPERIMENTAL RESULTS**

## CODE:

import pandas as pd

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, classification\_report from sklearn.model\_selection import train\_test\_split

# Train a Random Forest classifier on the training data

rfc = RandomForestClassifier(n\_estimators=100, max\_depth=10, random\_state=42) rfc.fit(X\_train, y\_train)

# Make predictions on the validation set y\_pred\_val = rfc.predict(X\_val)

# Evaluate the model performance on the validation set RF\_acc\_val = accuracy\_score(y\_val, y\_pred\_val) print('Validation Accuracy:', RF\_acc\_val) print(classification\_report(y\_val, y\_pred\_val))

# Make predictions on the testing set y\_pred\_test = rfc.predict(X\_test)

# Evaluate the model performance on the testing set RF\_acc\_test = accuracy\_score(y\_test, y\_pred\_test) print('Testing Accuracy:', RF\_acc\_test) print(classification\_report(y\_test, y\_pred\_test))

import pickle

# dump information to that file pickle.dump(rfc, open('model.pkl', 'wb'))

import pandas as pd import numpy as np

import joblib # or just import joblib if you're using a recent version

model\_path = 'model.pkl' # Update this path model = joblib.load(model\_path)

features\_no\_sepsis\_new = [

3, # Patient\_ID (unique identifier) 36,

7.2,

0,

0.03,

17,

28,

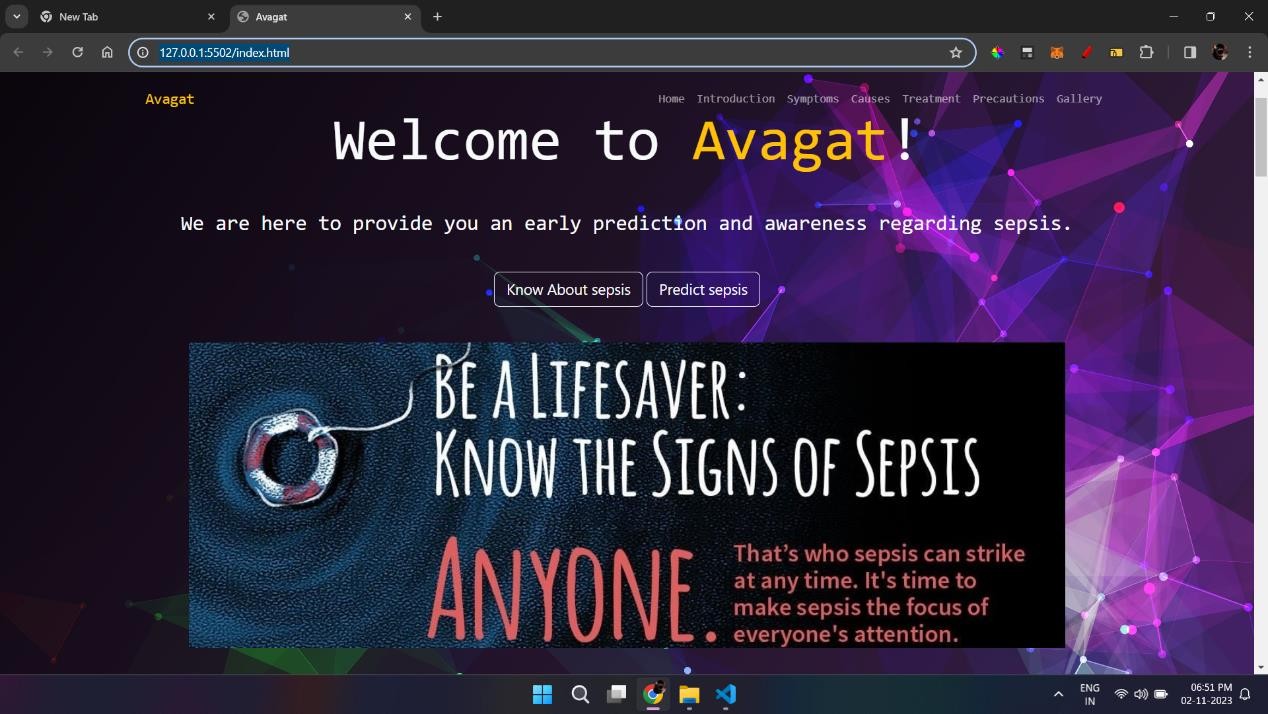
]

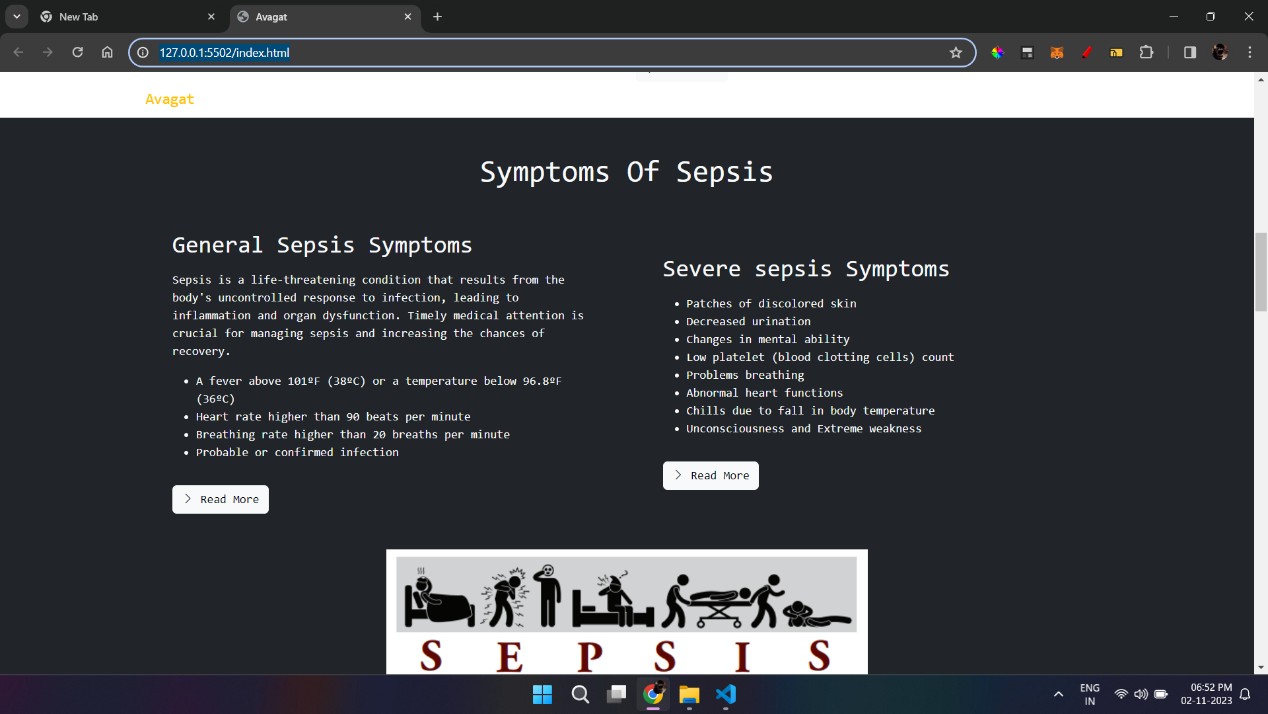
# Assuming the necessary imports and model loading have been done feature\_vector\_no\_sepsis\_new = np.array(features\_no\_sepsis\_new).reshape(1, -1)

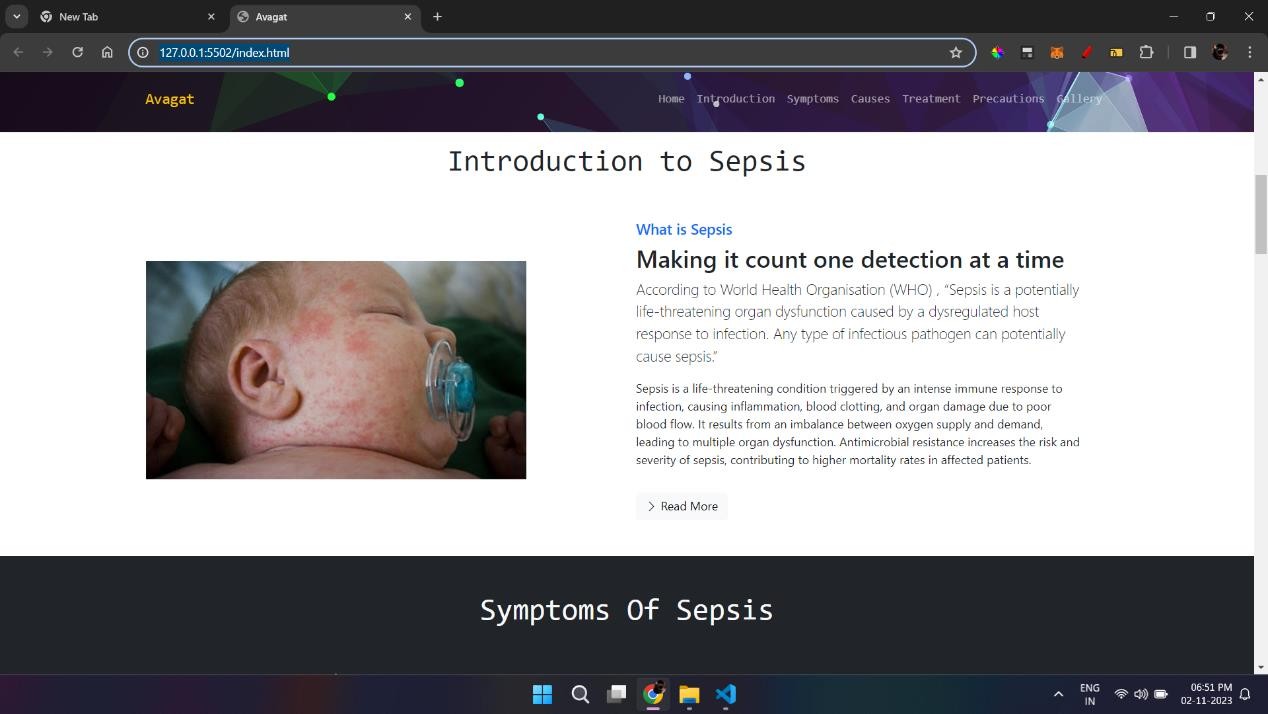
# Make a prediction using the new feature vector prediction\_new = model.predict(feature\_vector\_no\_sepsis\_new)

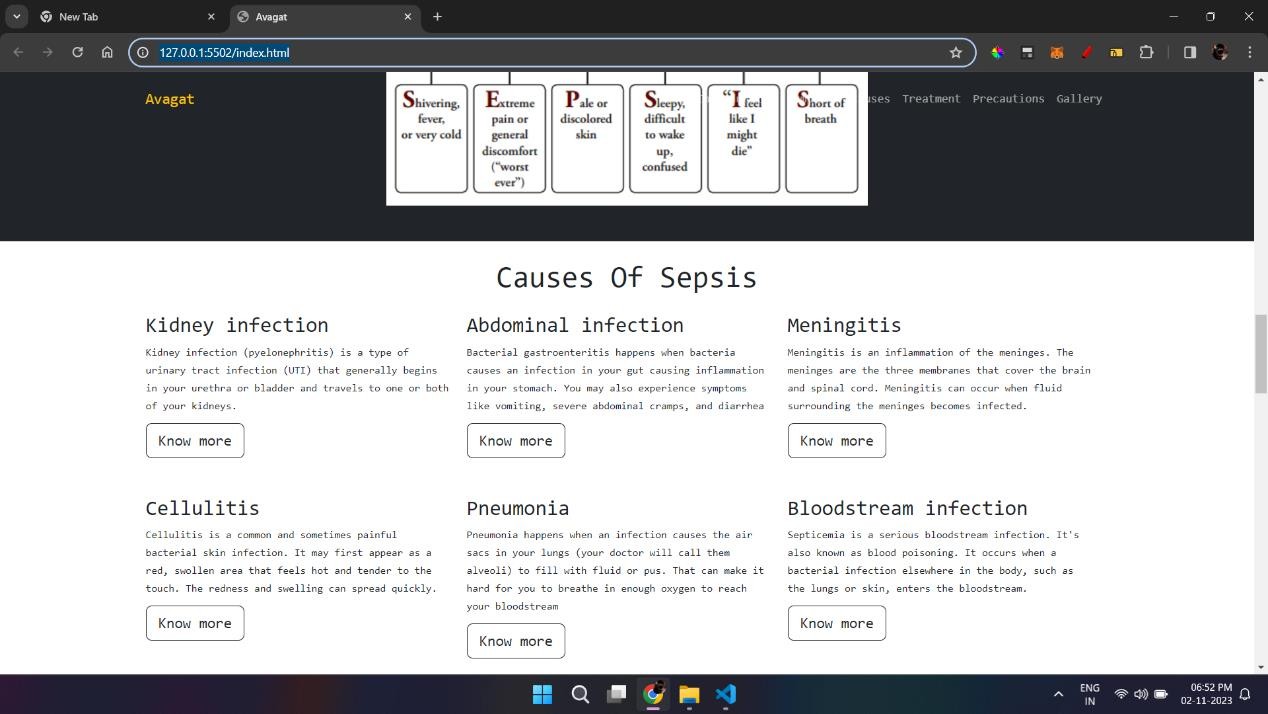
# Interpret the new prediction

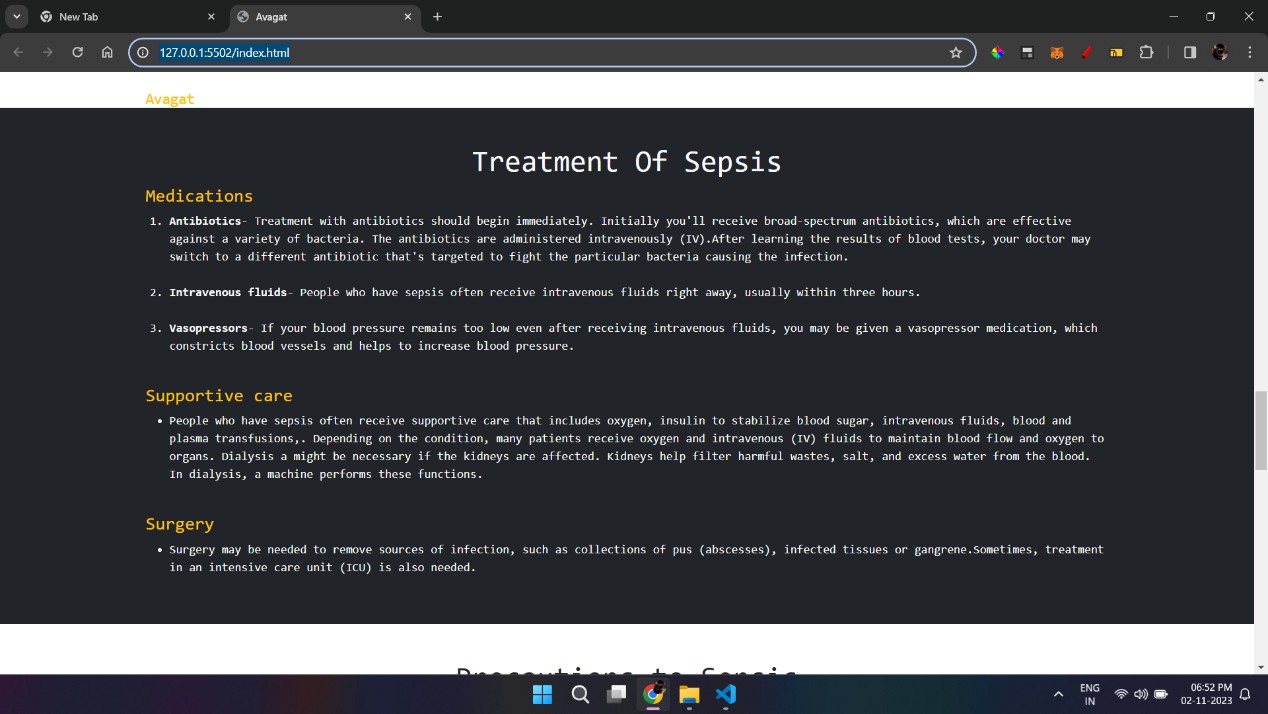
prediction\_result\_new = "Sepsis" if prediction\_new[0] == 1 else "No Sepsis" print(f"New Prediction: {prediction\_result\_new}")

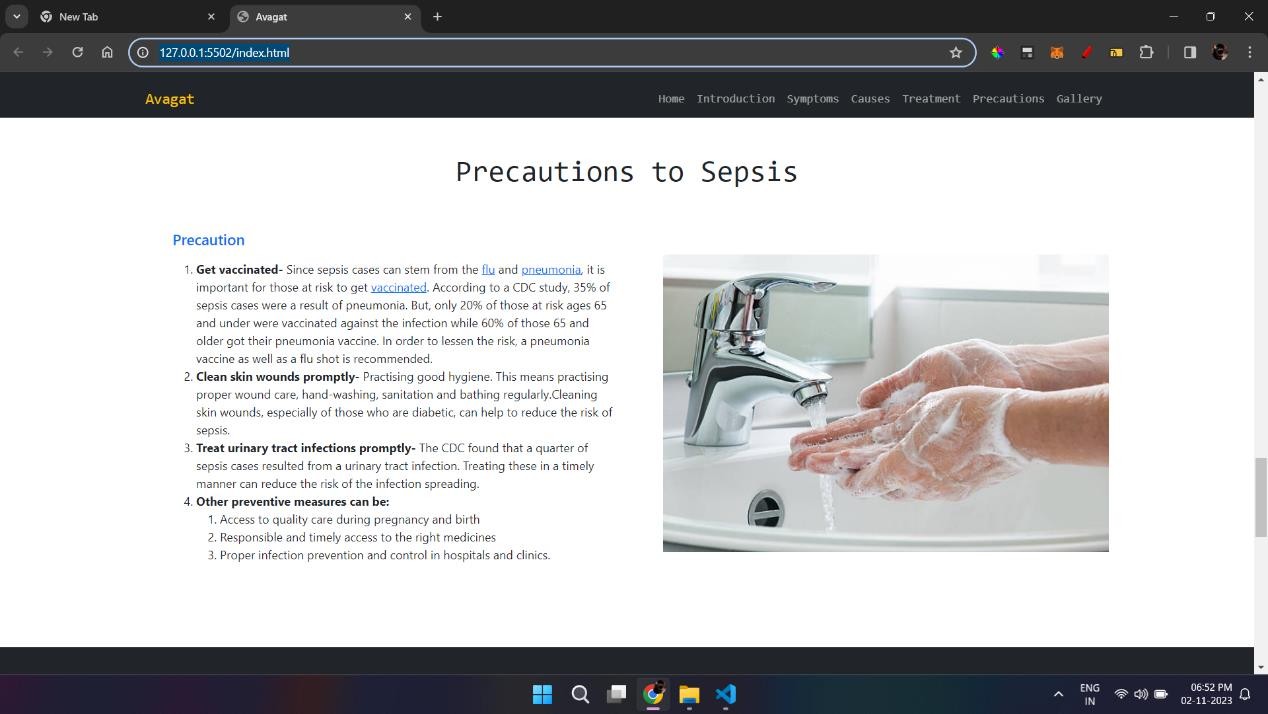


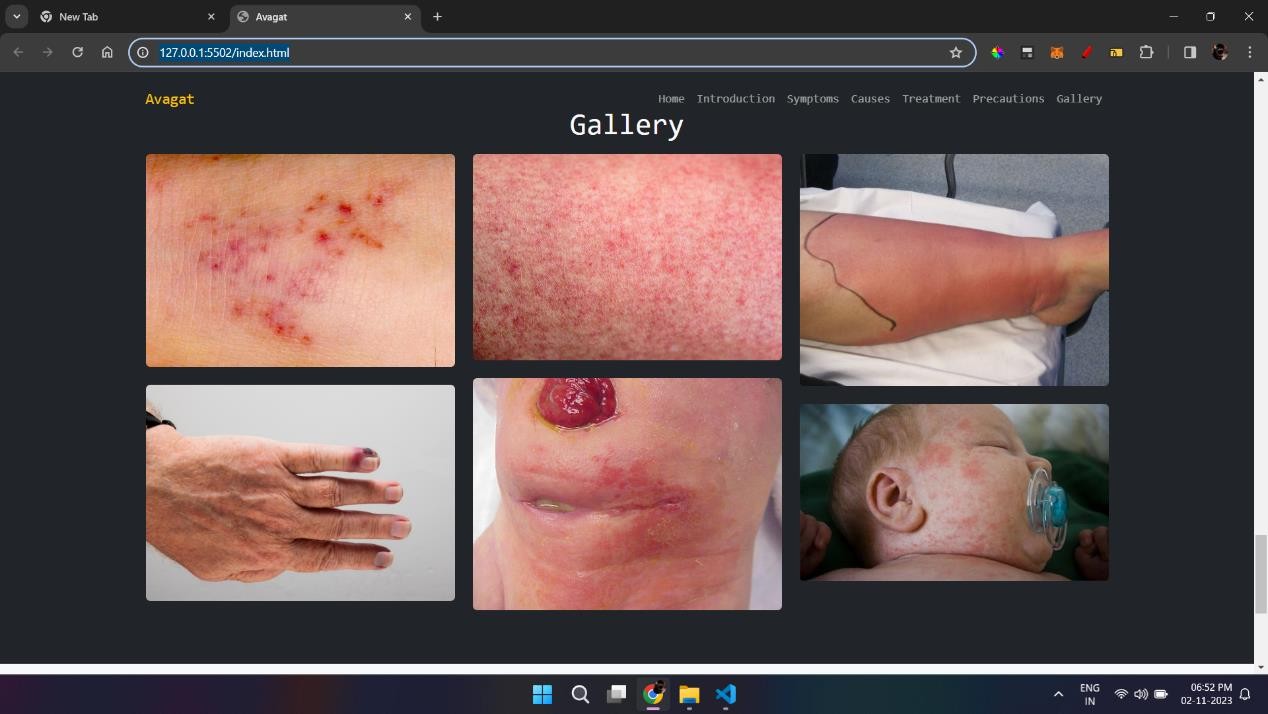


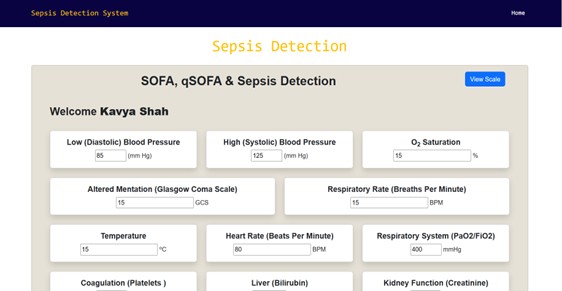




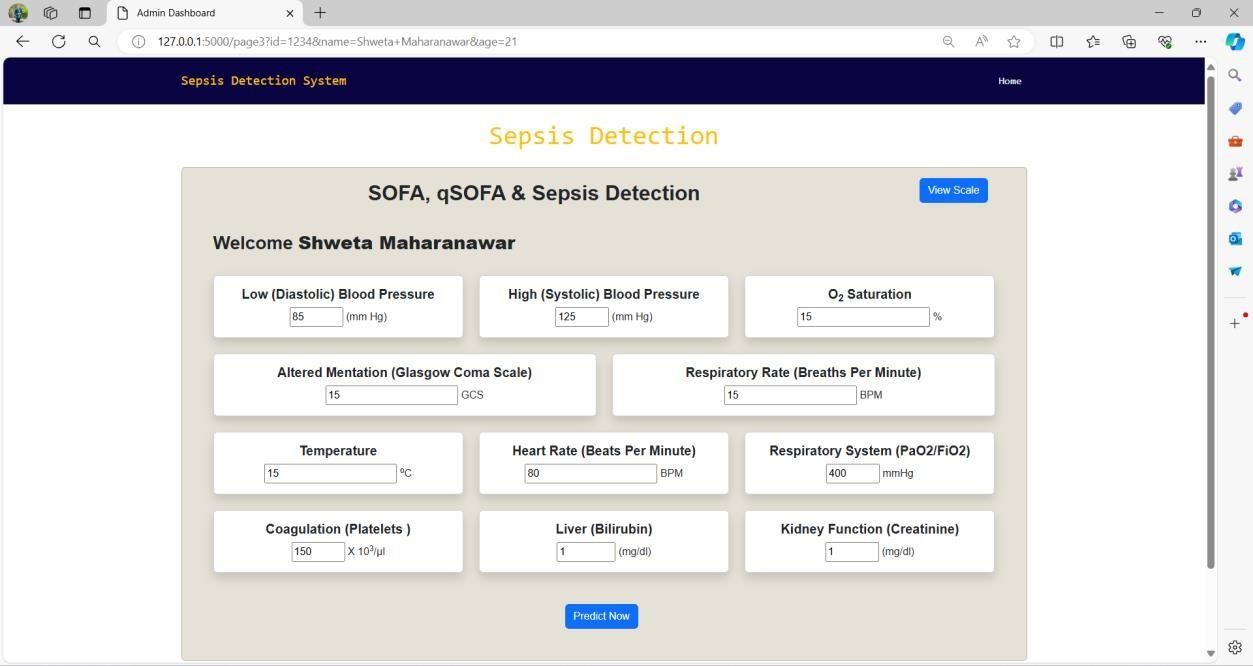




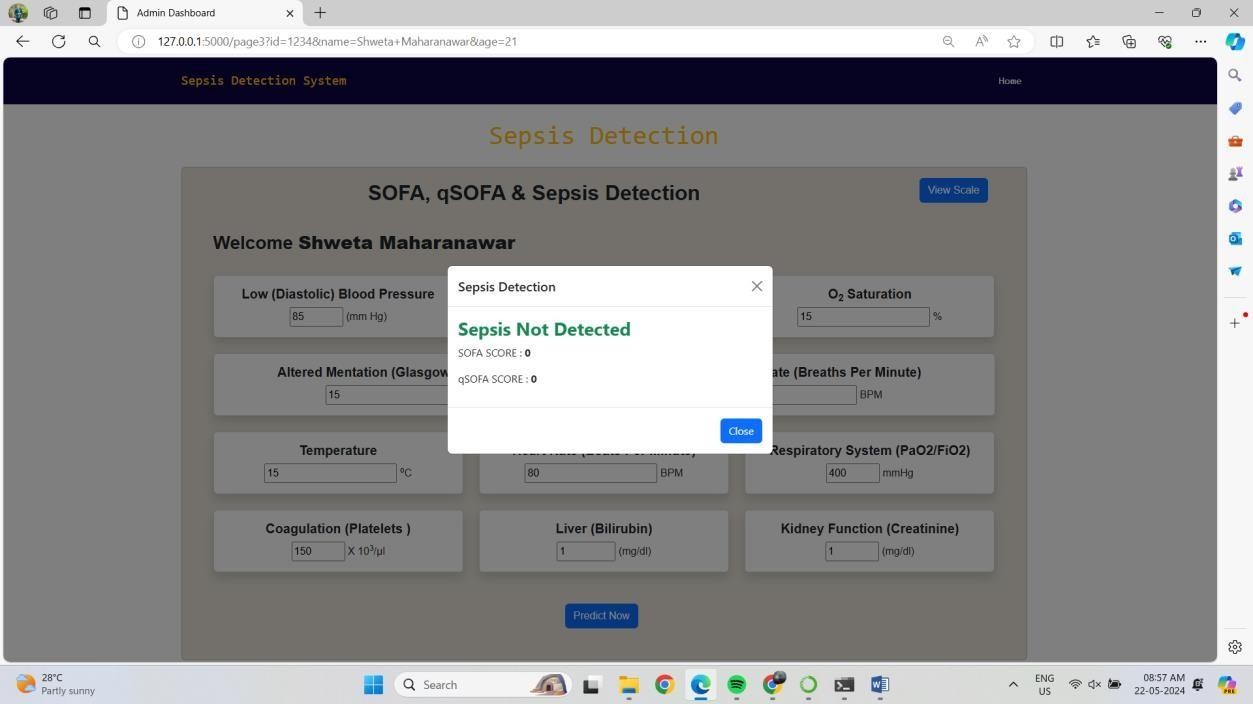




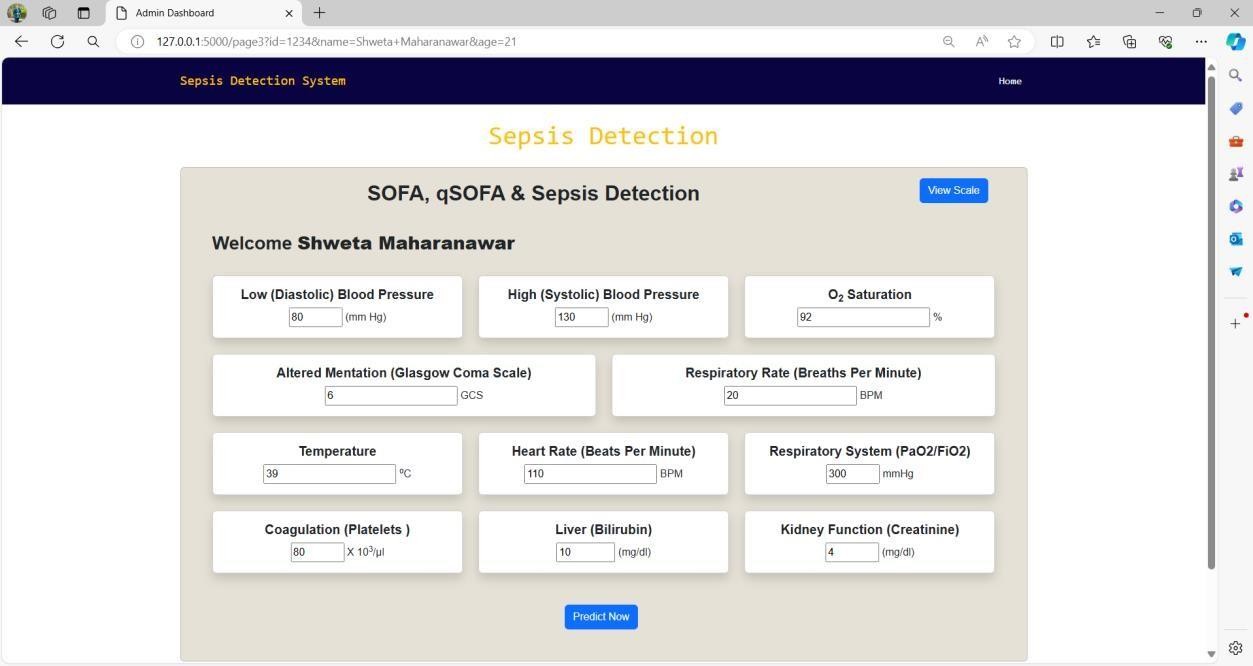
### User Input:



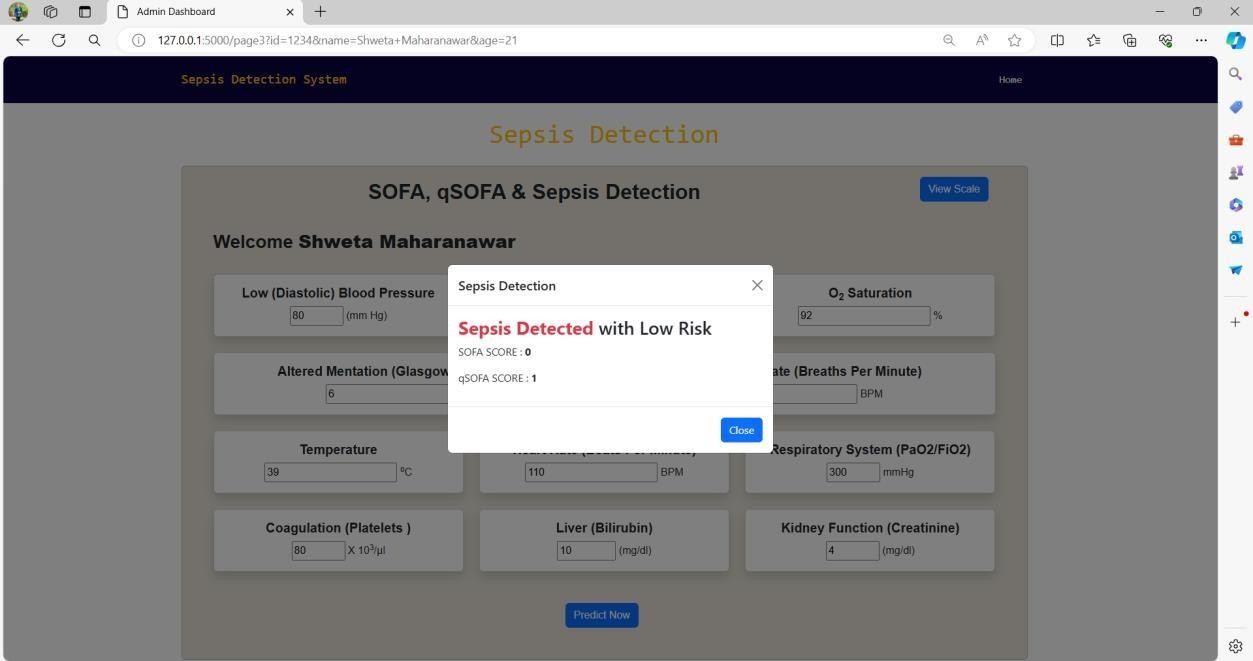
User Output:



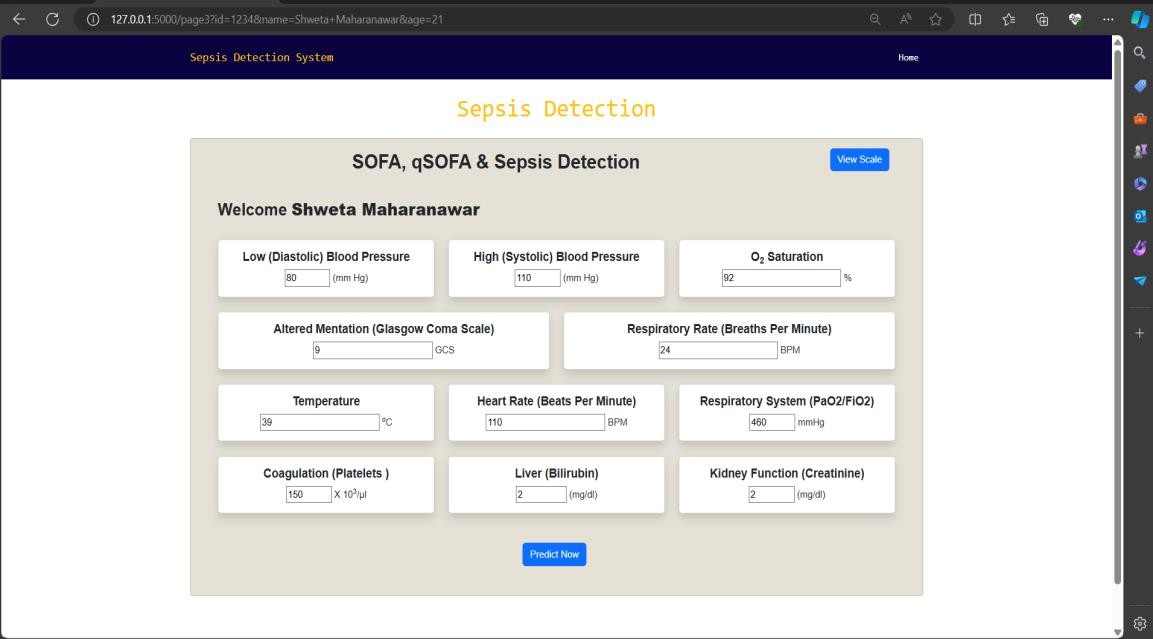
### User Input:



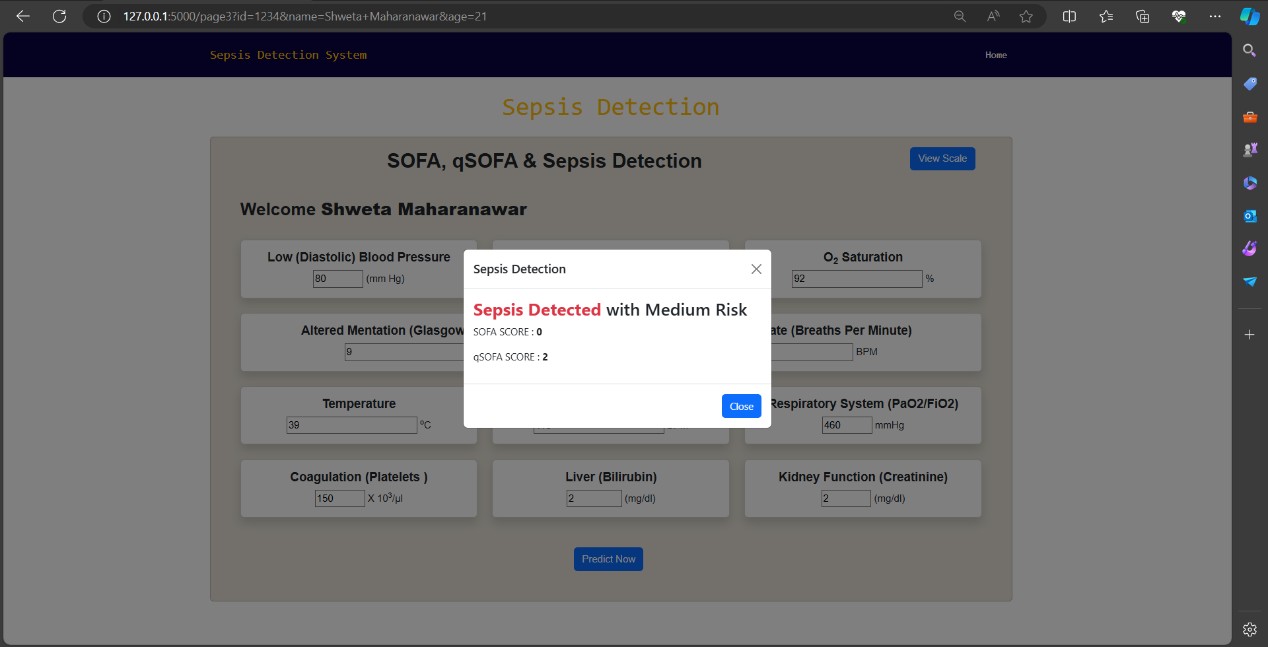
User Output:



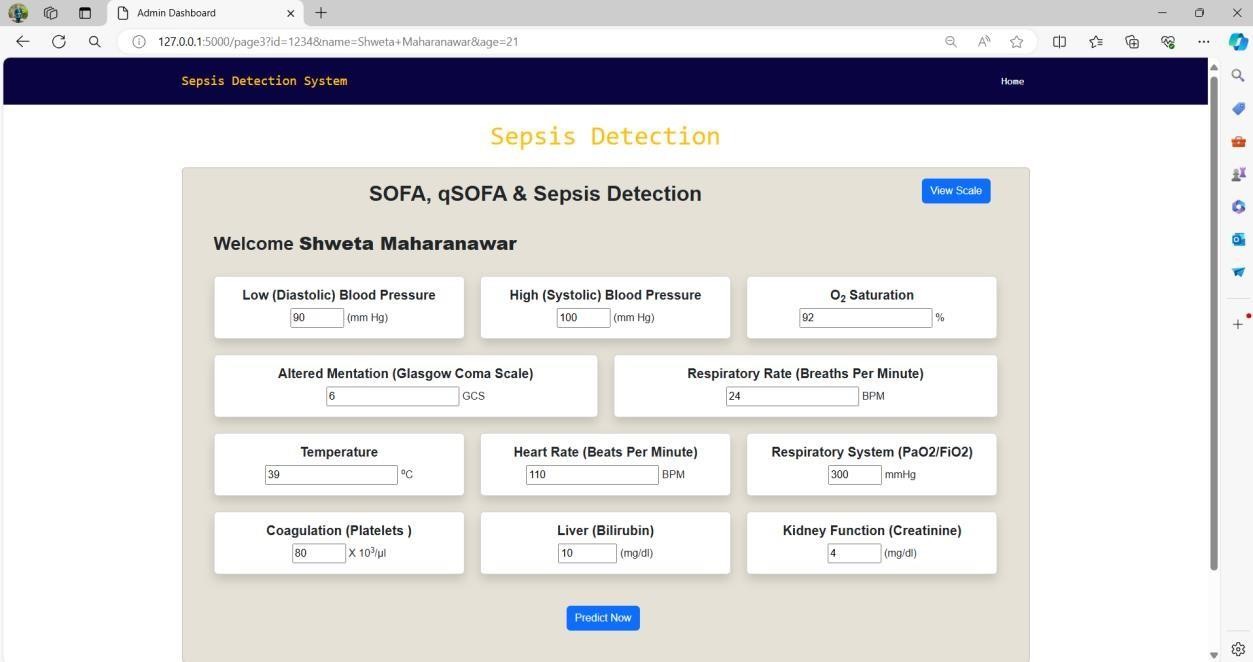
### User Input:



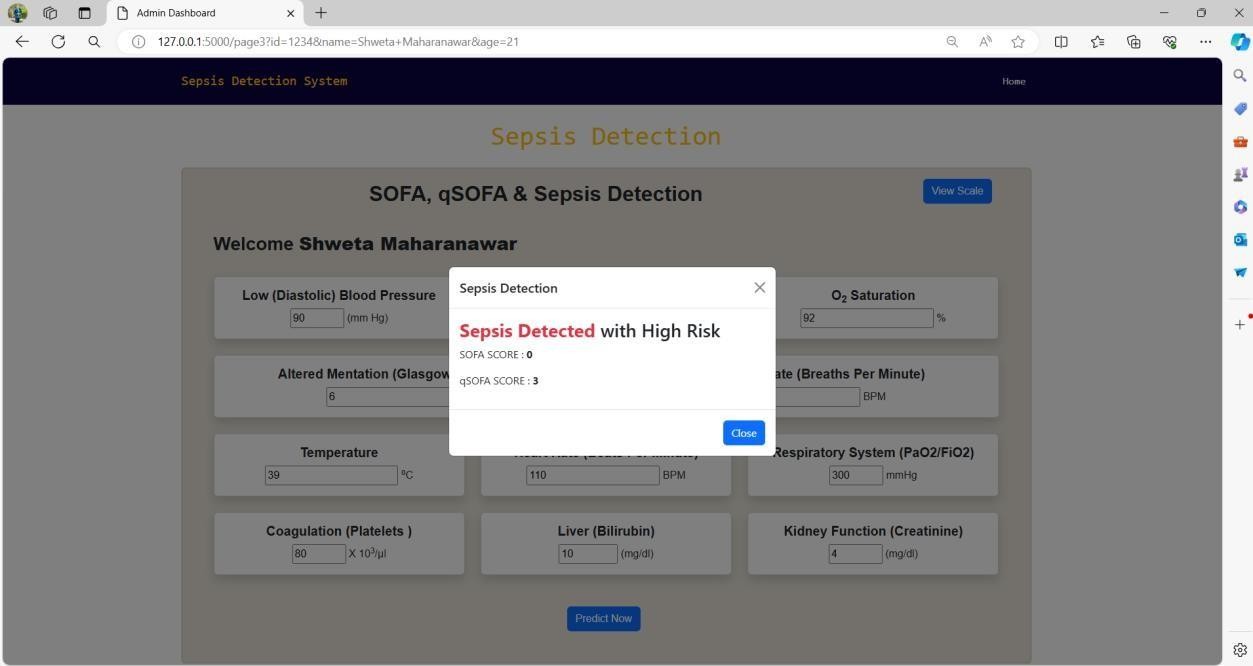
User Output:

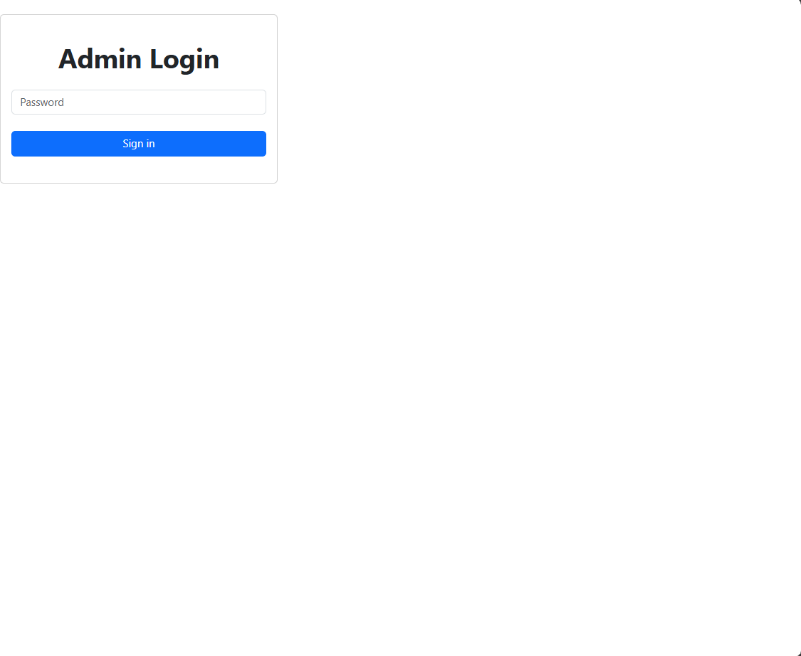


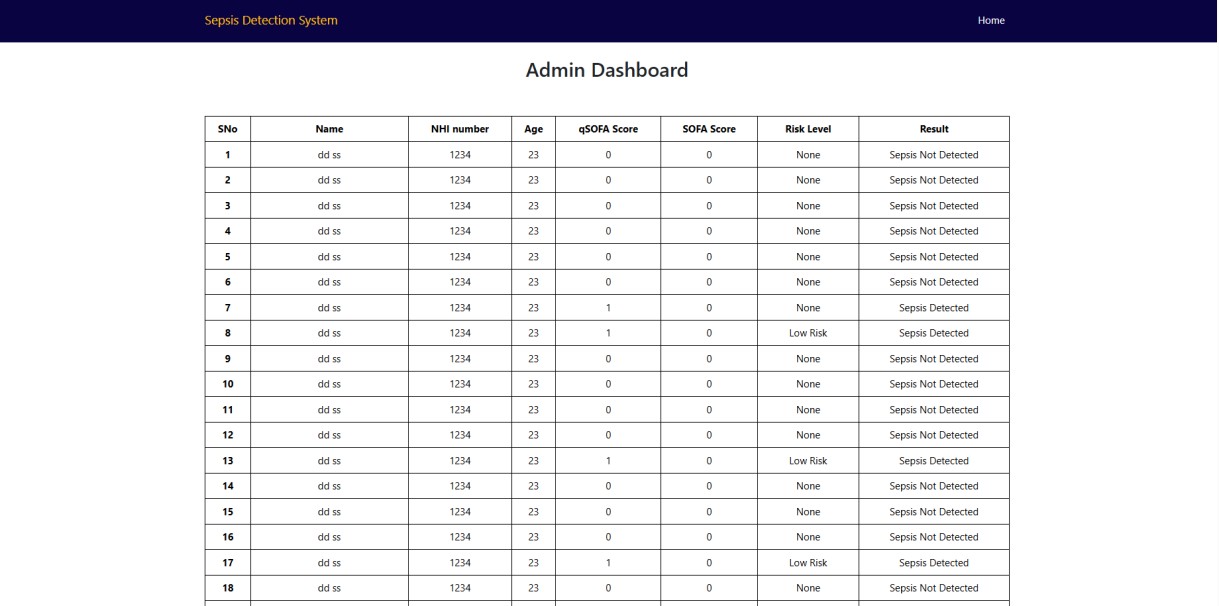
User Input:



User Output:







# CHAPTER 9 CONCLUSION

In conclusion, our project represents a pioneering endeavor that addresses the critical issues of sepsis awareness and early detection in intensive care units (ICUs). By integrating education, technology, and clinical insights, this project strives to create a holistic approach to sepsis management, with the ultimate goal of improving patient outcomes and reducing sepsis-related fatalities.

Through a user-friendly web interface and educational content, the project empowers individuals, both healthcare professionals and the general public, with the knowledge and tools to recognize sepsis symptoms early. This educational component, combined with the implementation of the Random Forest Regress or algorithm fine-tuned for accuracy, ensures rapid and accurate sepsis risk assessment.

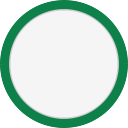
The project's significance is underscored by the potential to save lives and contribute to a healthier future. By raising sepsis awareness, fostering early detection, and ensuring the ethical and secure handling of healthcare data, the project sets a precedent for the successful integration of education and technology in healthcare practices.

As we look ahead, the project not only holds the promise of reducing the burden of sepsis but also serves as a model for addressing critical medical conditions through interdisciplinary collaboration. With its commitment to improving patient-centered care, the project stands as a testament to the potential of innovation in healthcare, setting a course toward a brighter and healthier future for all.

# BIBLIOGRAPHY

1. Anurag Shankar, Mufaddal Diwan, Snigdha Singh, Husain Nahrpurawala and Tanusri Bhowmick, “Early Prediction of Sepsis using Machine Learning,” IEEE 2021.
2. B. C. Srimedha, Rashmi Naveen Raj, And Veena Mayya, “A Comprehensive Machine Learning Based Pipeline for an Accurate Early Prediction of Sepsis in ICU”, IEEE access 2022.
3. Guilan Kong, Ke Lin, and Yonghua Hu,” Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU”, 2020, Published by BMC.
4. Yash Veer Singh, Pushpendra Singh, Shadab Khan, and Ram Sewak Singh, “A Machine Learning Model for Early Prediction and Detection of Sepsis in Intensive Care Unit Patients”, 2022, Published by Hindawi.
5. Longxiang Su, Zheng Xu, Fengxiang Chang, Yingying Ma, “Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models”, 2021, Published by Frontiers.
6. S. Kanaga Suba Raja, K. Valarmathi, S. Deepthi Sri, “Sepsis prediction using ensemble random forest”, 2022, Published by AIP Conference Proceedings.
7. S. Babu, Lalitha Anupama Annavarapu, Lakshmi Lahari Appala, “Sepsis Detection using Neural Networks”, 2022, Published by SMART GENCON.



**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 280 |
| **Characters** | 2166 |

**Content Checked For Plagiarism**

ABSTRACT

The "Accurate Prediction of Sepsis in ICU Patients" initiative is a comprehensive approach to addressing sepsis, a severe and often life-threatening condition frequently encountered in intensive care units (ICUs). This initiative combines an awareness campaign with advanced predictive modeling to improve early detection and intervention. The awareness campaign aims to educate the public and healthcare professionals about sepsis, emphasizing its severity, common symptoms—such as fever, increased heart rate, rapid breathing, confusion, and severe discomfort—and the critical need for prompt medical attention. By empowering individuals to recognize these signs early and seek immediate help, the campaign seeks to improve patient outcomes significantly.

Concurrently, the initiative employs advanced machine learning techniques, particularly the Random Forest algorithm, to develop a predictive model for sepsis. This involves comprehensive data preprocessing to ensure data quality, addressing class imbalances to avoid biased predictions, and incorporating clinical assessment tools like the Sequential Organ Failure Assessment (SOFA) and quick SOFA (qSOFA) criteria. These steps enhance the model's accuracy in identifying patients at high risk of sepsis, enabling timely and targeted interventions that can be life-saving.

A dedicated website supports this dual approach by serving as a central hub for sepsis education and the dissemination of the predictive model. It offers valuable resources, including educational materials and guidelines for recognizing sepsis symptoms, and provides healthcare professionals with access to the predictive model for clinical use.

In summary, the "Accurate Prediction of Sepsis in ICU Patients" initiative aims to raise sepsis awareness, improve predictive capabilities, and support healthcare professionals and the public in combating sepsis, contributing to a healthier and more informed future.

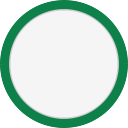
Keywords: Sepsis, Random Forest, Intensive care units (ICUs), Sequential Organ Failure Assessment (SOFA), quick SOFA (qSOFA).

**Matched Source**

No plagiarism found





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 575 |
| **Characters** | 4331 |

**Content Checked For Plagiarism**

CHAPTER 1

INTRODUCTION TO PROJECT

* 1. Introduction to Project

Sepsis, a life-threatening condition triggered by infections, remains a critical concern within the realm of intensive care units (ICUs). Our project " Accurate Prediction of Sepsis in ICU Patients" tackle this issue. This approach provides two

benefits: raising awareness about sepsis and predictive model for detection of sepsis.

At its core, prioritizes education and awareness. Our initiatives target both the general public and healthcare professionals because we understand how crucial knowledge is in combating sepsis. By educating people about the seriousness of sepsis and the critical need for early detection and intervention, we aim to enable individuals to identify its symptoms and seek

prompt medical care.

In parallel, the project leverages advanced machine learning techniques, specifically Random Forest, to construct a

predictive model for sepsis. The model's rigorous fine-tuning is aimed at achieving accurate identification of sepsis risk in ICU patients. This predictive system represents advancement in healthcare, offering early detection and improved outcome. Our data-driven approach encompasses comprehensive data preprocessing, addressing class imbalances within the dataset. Additionally, incorporating the Sequential Organ Failure Assessment (SOFA) score and the quick SOFA (qSOFA) criteria improves predictive accuracy by assessing organ failure trajectories and risk factors. Furthermore, the project features a dedicated website that serves as a vital platform for sepsis education and for distributing the predictive model to the medical community.

* 1. Aims/Motivation behind Project

The driving force behind the Project can be distilled into a few key motivations:

1. Sepsis Severity: The project is motivated by the gravity of sepsis as a life-threatening condition, especially within intensive care units (ICUs). Recognizing the severity of sepsis fuels our commitment to addressing this critical issue.
2. Early Detection: Early detection and intervention are pivotal in sepsis management. We are motivated to raise awareness and create a predictive model to enable early sepsis identification, ultimately saving lives.
3. Data-Driven Solutions: We are driven by the belief in data-driven solutions. Leveraging advanced machine learning

techniques, we seek to use data to enhance decision-making in healthcare.

1. Patient-Centered Care: Our core motivation is patient-centered care. We are dedicated to improving patient outcomes, reducing sepsis-related burdens, and contributing to a healthier future.
2. Interdisciplinary Approach: Recognizing the value of interdisciplinary methods, we are motivated to address sepsis comprehensively by combining education, technology, and clinical insights.
3. Raising Awareness: The first step in combating sepsis is awareness. We are motivated to spread knowledge about sepsis, ensuring both the public and healthcare professionals are equipped to recognize and respond to this condition.
4. Reducing Tragedies: Our ultimate motivation is to reduce sepsis-related tragedies. The knowledge that sepsis is a

leading cause of mortality in ICUs drives us to integrate clinical insights, including qSOFA, into our predictive model to make a tangible difference.

1.3 Overview of the Project

Our Project is executed following an Agile software development model. This model ensures flexibility, adaptability, and a continuous improvement approach. The project unfolds through iterative cycles, allowing for regular feedback and enhancements. Each cycle includes phases of requirement gathering, system design, implementation, testing, and deployment. Data preprocessing and modeling are central components, with Random Forest regression driving predictive capabilities. Integration of the Sequential Organ Failure Assessment (SOFA) score, including qSOFA criteria, enhances accuracy. Additionally, the project maintains a dedicated website for sepsis education and predictive model dissemination.

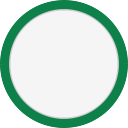
This approach enables the project to remain responsive to evolving needs, deliver incremental value, and align with its overarching goals effectively.

**Matched Source**

No plagiarism found





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 594 |
| **Characters** | 4222 |

**Content Checked For Plagiarism**

* 1. Need of the Project

Our Project addresses a pressing need within the healthcare landscape. Sepsis, a life-threatening condition, demands swift identification and intervention, particularly in intensive care units (ICUs) where patients are most vulnerable. The significance of this study lies in its potential to revolutionize patient care by bridging critical gaps. By fostering sepsis awareness among both the public and healthcare professionals, the project empowers individuals to recognize symptoms

early, ultimately saving lives. Furthermore, the development of an advanced predictive model for sepsis, integrated with the Sequential Organ Failure Assessment (SOFA) score and qSOFA criteria, ensures rapid and accurate risk assessment. This

innovation enhances patient outcomes and contributes to the reduction of sepsis-related fatalities in ICUs. The project's profound impact extends beyond the boundaries of sepsis, serving as a model for the successful integration of education, data-driven technology, and clinical insights in healthcare. Ultimately, the significance of this study lies in its potential to

create a healthier, more informed future, where sepsis-related tragedies are minimized, and the standards of patient care in ICUs are elevated.

* 1. Organization of the Project Report

1. Introduction to Project:

* Provide an overview of the project, introducing the significance of early sepsis detection in the ICU setting and the purpose of the Project.

1. Aims / Motivation behind Project:

* Explain the motivation for your project, which is to improve the accuracy of sepsis prediction using machine learning and enhance patient outcomes in ICUs.

1. Overview of the Project:

* Outline the main components of the project, including the utilization of SOFA and qSOFA scores, machine learning algorithms, and the goal of achieving early sepsis detection.

1. Need of the Project:

* Explore the reasons why accurate sepsis prediction in ICU patients is essential. Discuss the challenges and gaps in current sepsis detection methods that your project aims to address.

1. Literature Survey (minimum five papers):

* Provide a literature review section that covers relevant academic and research papers. Discuss the background, related work, and the problem statement in the context of sepsis prediction.

1. Features of the Project:

* Describe the core features of the project, focusing on the application of SOFA and qSOFA scores, machine learning models, and their integration into clinical practice.

1. Scope of the Project:

* Define the boundaries of your project, clarifying what aspects of sepsis detection in ICUs are covered. This sets

expectations for the reader.

1. Objectives of the Project:

* List the specific goals and objectives of your project, highlighting what you intend to achieve with regard to sepsis prediction.

1. Constraints of the Project:

* Address any limitations or constraints that may have influenced the development of the project, such as data availability or model performance.

1. Project Requirements:

* Outline the hardware and software requirements necessary to implement and run project. This section explains what's needed for the platform to function effectively.

1. System Analysis of Proposed Architecture:

* Delve into the technical aspects of your project, including system architecture, data flow diagrams, and any relevant UML diagrams. This provides a technical understanding of how the platform operates.

1. System Implementation:

* Discuss the technical implementation of your sepsis prediction system, including the use of machine learning algorithms, data preprocessing, and the integration of SOFA and qSOFA scores.

1. Project Plan:

* Present the project's timeline, milestones, and development plans. This section outlines how the project will be executed.

1. Conclusion:

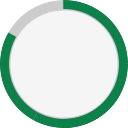
* Summarize the project's key takeaways, achievements, and its potential impact on improving sepsis detection in ICU patients. Reflect on the project's journey and outcomes.

**Matched Source**

No plagiarism found





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 17%  Plagiarised | 83%  Unique | **Date** | 2024-05-22 |
| **Words** | 457 |
| **Characters** | 3436 |

**Content Checked For Plagiarism**

CHAPTER 2 LITERATURE SURVEY

* 1. Introduction

This literature survey provides an overview of relevant research in the field of sepsis prediction, emphasizing the importance of early diagnosis and the application of machine learning techniques. By reviewing recent studies, we aim to establish a contextual framework for our project. These studies highlight the significance of predictive models, such as Random Forest, Logistic Regression, and Gradient Boosting, in predicting sepsis and in-hospital mortality. Through this literature survey, we aim to gain insights from existing work and inform the direction of our research.

* 1. Related Work

Title Author Publication Remark

Early Prediction of Sepsis using Machine Learning Anurag Shankar, Mufaddal Diwan, Snigdha Singh, Husain Nahrpurawala and Tanusri Bhowmick

IEEE, 2021 The paper introduces the importance of the SOFA score and the challenges in timely diagnosis, serving as a basis for our project.

A Comprehensive Machine Learning Based Pipeline for an Accurate Early Prediction of Sepsis in ICU B. C. SRIMEDHA, RASHMI NAVEEN RAJ, AND VEENA MAYYA IEEE Access, 2022 This research by investigating four prediction algorithms, including Random Forest, Logistic Regression, Gradient Boosting, and Decision Tree, and examining the impact of various imputation techniques,

Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU Guilan Kong, Ke Lin and Yonghua Hu 2020, Published by

BMC. This study focuses on leveraging machine learning techniques to predict the in-hospital mortality of sepsis patients

in the ICU.

Machine learning models, including the least absolute shrinkage and selection operator (LASSO), random forest (RF), gradient boosting machine (GBM), and traditional logistic regression (LR), were developed for prediction.

A Machine Learning Model for Early Prediction and Detection of Sepsis in Intensive Care Unit Patients Yash Veer Singh, Pushpendra Singh, Shadab Khan, and Ram Sewak Singh 2022, Published by Hindawi. This paper introduces a machine learning model for early sepsis prediction in ICU patients, leveraging data from clinical laboratory values and vital signs. Various models, including SVM, RF, NB, LR, and XGBoost, are examined and compared, with the proposed ensemble method showing the most promising results in terms of classification performance and prognosis improvement.

Chang, Yingying Ma, Shengjun Liu, Huizhen Jiang, Hao Wang, Dongkai Li, Huan Chen, Xiang Zhou, Na Hong, Weiguo Zhu, and Yun Long 28 June 2021, Published by Frontiers in Medicine This study centers around harnessing machine learning techniques to anticipate in-hospital mortality among sepsis patients in the ICU. Several machine learning models were constructed for this purpose, encompassing the least absolute shrinkage and selection operator (LASSO), random forest (RF), gradient boosting machine (GBM), and the conventional logistic regression (LR).

Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis

3.0 by Machine Learning Models Longxiang Su, Zheng Xu, Fengxiang

**Matched Source**

**Similarity** 20%

**Title**:A Comprehensive Machine Learning Based Pipeline for an ...

A Comprehensive Machine Learning Based Pipeline for an Accurate Early Prediction of Sepsis in ICU ... Authors: B. C. Srimedha, Rashmi Naveen Raj, Veena Mayya ...

<https://ouci.dntb.gov.ua/en/works/4wdvmdGl/>

**Similarity** 6%

**Title**:Using machine learning methods to predict in-hospital mortality ...

We employed the least absolute shrinkage and selection operator (LASSO), random forest (RF), gradient boosting machine (GBM) and the traditional logistic ...

<https://www.springermedizin.de/using-machine-learning-methods-to-predict-in-hospital-mortality-/18442368>

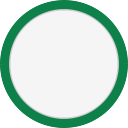
**Similarity** 6%

**Title**:Clinical Application of Artificial Intelligence in Emergency ...

<https://books.google.com/books?id=JXRcEAAAQBAJ>





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 528 |
| **Characters** | 4057 |

**Content Checked For Plagiarism**

CHAPTER 3

PROBLEM STATEMENT

To develop a predictive model for early detection of sepsis in ICU patients and to create a website for generating a website for sepsis awareness containing information about sepsis.

* 1. Features of the Project:
* Sepsis Awareness Campaign: The project includes a dedicated website and educational resources to raise awareness about sepsis, its symptoms, and the importance of early detection.
* Educational Content: Informative materials simplify complex medical information, making it accessible to the general

public and healthcare professionals, empowering them to recognize sepsis symptoms.

* Advanced Predictive Modeling: The project leverages Random Forest regression for the development of a predictive model that accurately identifies sepsis risk in ICU patients.
* Integration of qSOFA Criteria: The predictive model integrates the quick Sequential Organ Failure Assessment (qSOFA) criteria, enabling rapid risk assessment and intervention.
* Data Preprocessing: Comprehensive data preprocessing techniques are applied to address class imbalances in the dataset, enhancing the model's predictive accuracy.
* Interdisciplinary Approach: The project combines education, data-driven technology, and clinical insights, ensuring a holistic approach to sepsis management.
* Digital Platform: The project maintains a dedicated website as a central platform for sepsis education and the

dissemination of the predictive model to the medical community.

* Continuous Improvement: An agile development model allows for flexibility and ongoing enhancements to meet evolving needs effectively.
* Data-Driven Insights: The project ensures that data-driven insights are used to inform clinical decisions and enhance early intervention strategies.
  1. Scope of the Project:

The scope of our project encompasses creating awareness, implementing advanced predictive modeling for sepsis detection, and fostering a holistic approach to sepsis management. It aims to reduce sepsis-related mortality rates in ICUs, improve patient outcomes, and serve as a model for integrating education and technology in healthcare practices.

* 1. Objectives of the Project:
* To educate the public and healthcare professionals about sepsis, its severity, and the importance of early recognition.
* To develop and implement an advanced predictive model for sepsis to enable early detection and intervention in intensive care units (ICUs).
* To improve patient outcomes by minimizing sepsis-related fatalities through early intervention and rapid risk assessment.
* To utilize data-driven insights to inform clinical decisions and enhance the quality of patient care in ICUs.
* To combine education, technology, and clinical insights to create a holistic approach to sepsis management.
* To contribute to a reduction in the burden of sepsis by fostering proactive healthcare practices.
* To serve as a model for successfully integrating education and technology into healthcare practices, potentially impacting other critical medical conditions.
  1. Constraints of the Project:
* Data Availability: The project relies on the availability of comprehensive and reliable healthcare data for the development of the predictive model, which can sometimes be limited.
* Technological Infrastructure: Access to suitable technological infrastructure and resources is crucial for implementing the predictive model and maintaining the project's digital platform.
* Clinical Adoption: The successful adoption of the predictive model within clinical settings may face resistance or

challenges, requiring careful integration strategies.

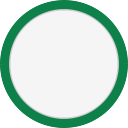
* Data Privacy and Ethics: Adhering to data privacy and ethical considerations in healthcare data usage is a paramount constraint that demands rigorous compliance.

**Matched Source**

No plagiarism found





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 280 |
| **Characters** | 2130 |

**Content Checked For Plagiarism**

CHAPTER 4

PROJECT REQUIREMENTS

Project Requirements are as follows:

* Data Set: Access to a comprehensive healthcare data set, sourced from Kaggle, is crucial for the development and training of the predictive model.
* Algorithm: Implementing the Random Forest predictive model, ensuring its accuracy.
* HTML and CSS: Develop a user-friendly web interface using HTML and CSS to provide educational resources, awareness materials, and model access to users.
* Flask Framework: Utilize the Flask web framework to build the project's digital platform, integrating the educational

content, predictive model, and awareness campaigns.

* Database: Set up a database to store and manage healthcare data required for the predictive model, ensuring data security, privacy, and efficient retrieval.
* Technology Infrastructure: Ensure access to the necessary technological infrastructure, including server resources, for hosting the project's website and predictive model.
* Budget and Funding: Secure adequate funding to sustain the project, covering the costs associated with awareness campaigns, server hosting, and development efforts.
* Data Privacy and Ethics: Adhere to stringent data privacy and ethical standards when handling sensitive healthcare data within the project.
* Flexibility: Maintain a flexible project approach to accommodate the evolving landscape of medical practices and

technological advancements.

By fulfilling these requirements, the project can effectively bridge the gap in sepsis awareness and early detection while contributing to improved patient care in ICUs.

4.1 Hardware and Software Requirements

* Hardware

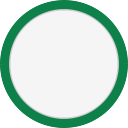
1. Computer(s)
2. System: Intel I3 processor or above
3. RAM: 4GB or above

* Software

1. Operating System: Windows 10 or any open-source OS
2. Programming Languages: Python for machine learning, data development.
3. Development Tools: Visual Studio Code, Anaconda

* Other equipment



**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 421 |
| **Characters** | 3171 |

**Content Checked For Plagiarism**

CHAPTER 5

SYSTEM ANALYSIS OF PROPOSED ARCHITECTURE

* 1. Proposed System Architecture

Fig 5.1 Proposed System Architecture

The system architecture for the Accurate Prediction of Sepsis in ICU Patients is designed to integrate technology, education, and clinical insights to address the issue of sepsis in intensive care units (ICUs).

1. Input Data:

* The architecture comprises a set of vital signs and clinical parameters from ICU patients as input data. Specifically, we focus on:
* Low (Diastolic) Blood Pressure
* High (Diastolic) Blood Pressure
* O2 Saturation
* Altered Mentation (Glasgow Coma Scale)
* Respiratory Rate (Breaths per minute)
* Temperature
* Heart Rate (Beats per minute)
* Respiratory Rate (PaO2/FiO2)
* Liver (Bilirubin)
* Kidney Function (Creatinine)

1. SOFA Score Calculation:

* The SOFA score is a comprehensive assessment that evaluates the patient's condition across various organ systems. This includes:
* Respiratory System
* Liver
* Kidney
* Machine learning algorithms are employed to calculate the SOFA score by considering the values of these organ systems.

1. qSOFA Score Calculation:

* In contrast to SOFA, the qSOFA score is a quicker assessment.
* Low Blood Pressure (SBP <= 100mmHg)
* Respiratory Rate (>= 22 breaths /min)
* Altered Mentation (GCS <= 14)
* We use machine learning techniques to calculate the qSOFA score, which allows quick predictions of sepsis risk.

1. Prediction Outcome:

* Both the SOFA and qSOFA scores play a crucial role in our sepsis prediction. These scores are used as features to train machine learning models. This integrates the scores with other clinical parameters to enhance accuracy.

1. Machine Learning Algorithms:

* While we use machine learning algorithms to calculate the scores, we also employ additional machine learning models for sepsis prediction. These models are fine-tuned for accuracy and will be chosen based on their performance. Future sections of the project report will provide detailed information on the specific machine learning algorithms used. Our architecture allows us to leverage clinical data, vital signs, and organ system assessments to make informed predictions about a patient's risk of developing sepsis. By using SOFA and qSOFA scores as integral components, we ensure a comprehensive and timely approach to sepsis prediction in ICU patients.
  1. High Level Design of the Project
     1. Data Flow Diagrams Level 0 DFD

Fig 5.2 Level 0 DFD

Level 1 DFD

Fig 5.3 Level 1 DFD

Level 2 DFD

Fig 5.4 Level 2 DFD

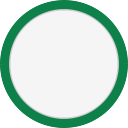
* + 1. UML Diagrams Use Case Diagram

**Matched Source**

No plagiarism found





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 630 |
| **Characters** | 4040 |

**Content Checked For Plagiarism**

CHAPTER 6

SYSTEM IMPLEMENTATION

* 1. Algorithm Style
     1. Random Forest

1. Bootstrap Aggregating (Bagging)

Each decision tree in a Random Forest is built from a sample drawn with replacement (i.e., a bootstrap sample) from the training set. For example, if our training set has 1000 rows, each decision tree gets a sample of 1000 rows drawn randomly from the training set, allowing the same row to appear multiple times in each sample.

1. Feature Randomness

When building each tree, each time a split in a tree is considered, a random subset of the features is chosen as split candidates from the full set of features. This is called the random subspace method. The purpose is to decrease the correlation between the trees in the forest. Typically, for a classification problem, the square root of the total number of features is used, whereas for algorithm, one-third of the total number of features is chosen.

1. How It Works

Each tree gives a prediction in the case of regression, the average of all the trees’ predictions is the final prediction. For classification, the prediction is the class with the majority votes across all trees.

Internal working of algorithm:

Step 1: Bootstrapping the Data

For each tree in the forest, algorithm starts by performing bootstrapping, which is a resampling technique. In this process, random samples of the training dataset are drawn with replacement. This means some observations may be repeated in each bootstrap sample, while others may be left out. Each bootstrap sample is used to train a separate decision tree, which helps ensure that the trees in the forest are de-correlated.

Step 2: Building Decision Trees

Random Forests build decision trees slightly differently from the usual method:

1. Random Feature Selection: When creating a split at a node in a decision tree, Random Forest randomly selects a subset of the features rather than considering every possible feature. The size of this subset can be controlled by

parameters such as max\_features in scikit-learn. For classification tasks, this is typically the square root of the total number

of features, and for regression, it’s about one-third. This randomness helps to make the model more robust and prevents overfitting.

1. Node Splitting: Each tree grows during the training process by splitting nodes based on the selected features. The

splits are chosen to best separate the classes in classification or minimize the variance in regression. This is typically done using measures such as Gini impurity or entropy for classification and mean squared error for regression.

3. Tree Depth: Trees in algorithm are generally grown to their full depth, but parameters such as max\_depth can limit

this. Full trees can mean that each tree can potentially overfit the data it was trained on, but when aggregated, the forest as a whole does not overfit as much.

Step 3: Aggregation of Predictions After all trees are trained:

* Classification: Each tree in the forest outputs a class prediction. The final prediction of the Random Forest is the class

that receives the majority vote among all the trees.

* Regression: Each tree predicts a continuous value. The final prediction of the Random Forest is typically the average of these predictions.

Step 4: Handling Overfitting and Variance

It handles overfitting through two mechanisms:

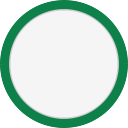
1. Bootstrap Aggregating (Bagging): Since each tree is trained on a different sample of data, the variance of the model is reduced. Even if some trees overfit their samples, the averaging makes the overall model less prone to overfitting.
2. Feature Randomness: By forcing each split to consider only a subset of the features, the trees are made less correlated with each other. Less correlation among trees in the model means a lower chance that they will all repeat the same errors, thus reducing the model variance further.

**Matched Source**

No plagiarism found





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 385 |
| **Characters** | 2875 |

**Content Checked For Plagiarism**

In our project, Random Forest can do various task like:

* Predictive Modelling: It is a supervised learning and bagging technique that uses an ensemble learning method for prediction in machine learning. The trees here run in parallel, meaning there is no interaction between these trees while building the trees. Random Forest can build a predictive model aimed at determining the likelihood of a patient developing sepsis during their ICU stay, based on diverse patient-related factors and clinical data.
* Feature Selection: It can help identify which patient variables or features are most relevant for predicting sepsis. This involves selecting and weighting various clinical and laboratory measurements, vital signs, and other patient data that are considered potential risk factors for sepsis.
* Risk Stratification: Random forest regression can help in stratifying patients into different risk categories based on their calculated sepsis risk scores. This can assist healthcare providers in prioritizing care and interventions for patients at higher risk of sepsis.
  1. Description of Detailed Methodologies Data Acquisition and Preprocessing:

It begins by collecting healthcare data from a dataset sourced from Kaggle. This data is then preprocessed to handle class imbalances and ensure its suitability for predictive modeling.

Predictive Modeling:

We've fine-tuned algorithm to improve early sepsis detection accuracy. The predictive model developed using Random Forest has yielded results in forecasting the likelihood of patients developing sepsis during their ICU stay. With an

impressive accuracy rate of 85%, this model represents a significant advancement in sepsis prediction, providing healthcare professionals with valuable insights for proactive intervention and patient management.

Web Interface:

It maintains a user-friendly web interface built with HTML and CSS, providing access to educational content and the predictive model. Users, including the general public and healthcare professionals, can access information on sepsis awareness, symptoms, and risk assessment.

Flask Web Framework:

It serves as the backbone of the digital platform. It handles user requests, routes them to the components, and ensures better communication between the user interface and the server.

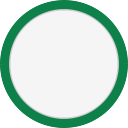
Interdisciplinary Collaboration:

Effective collaboration between healthcare professionals, data scientists, web developers, and educators is pivotal. It ensures that the project combines expertise in medical practices, data analysis, web development, and educational content.

Security and Privacy Compliance:

The system adheres to stringent security and data privacy standards, safeguarding sensitive healthcare information and ensuring ethical data usage.



**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 151 |
| **Characters** | 944 |

**Content Checked For Plagiarism**

CHAPTER 7

TEST CASES

No. Functional Test Cases Actual Output Expected Output Test Status

1

2

3

Pass 4

Pass

5

6

7

Verify if the User is able to see the complete UI of the system.

The UI is visible. The UI is visible. Pass

Verify if sepsis s detected with low risk. Sepsis detected with low risk. Sepsis detected with low risk. Pass

Verify if sepsis s detected with medium risk. Sepsis detected with medium risk. Sepsis detected with medium risk.

Verify if sepsis s detected with high risk.

Sepsis detected with high risk.

Sepsis detected with high risk.

Verify Admin login working properly. Works properly. Works properly.

Pass

Verify user data is saved in database.

Data saved successfully. Data saved successfully. Pass

Verify if the training model is giving good accuracy or not Showed Accuracy of 85%

Should show accuracy of

minimum 80% Pass

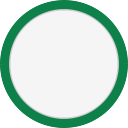
Table 7.1. Test case

**Matched Source**

No plagiarism found





**PLAGIARISM SCAN REPORT**

|  |  |  |  |
| --- | --- | --- | --- |
| 0%  Plagiarised | 100%  Unique | **Date** | 2024-05-22 |
| **Words** | 222 |
| **Characters** | 1620 |

**Content Checked For Plagiarism**

CHAPTER 9 CONCLUSION

In conclusion, our project represents a pioneering endeavor that addresses the critical issues of sepsis awareness and early detection in intensive care units (ICUs). By integrating education, technology, and clinical insights, this project strives to create a holistic approach to sepsis management, with the ultimate goal of improving patient outcomes and reducing sepsis-related fatalities.

Through a user-friendly web interface and educational content, the project empowers individuals, both healthcare

professionals and the general public, with the knowledge and tools to recognize sepsis symptoms early. This educational component, combined with the implementation of the Random Forest Regress or algorithm fine-tuned for accuracy, ensures rapid and accurate sepsis risk assessment.

The project's significance is underscored by the potential to save lives and contribute to a healthier future. By raising sepsis awareness, fostering early detection, and ensuring the ethical and secure handling of healthcare data, the project sets a

precedent for the successful integration of education and technology in healthcare practices.

As we look ahead, the project not only holds the promise of reducing the burden of sepsis but also serves as a model for addressing critical medical conditions through interdisciplinary collaboration. With its commitment to improving patient- centered care, the project stands as a testament to the potential of innovation in healthcare, setting a course toward a

brighter and healthier future for all.

**Matched Source**

No plagiarism found



***2021 2nd Global Conference for Advancement in Technology (GCAT) Bangalore, India. Oct 1-3, 2021***

Early Sepsis Detection using Machine Learning and Neural Networks

Nemil Shah\*

*Student, Computer Engineering DJ Sanghvi College of Engineering*

Mumbai, India [shahnemil103@gmail.com](mailto:shahnemil103@gmail.com)

Rishi Desai\*

*Student, Computer Engineering DJ Sanghvi College of Engineering*

Mumbai, India [rishindesai17@gmail.com](mailto:rishindesai17@gmail.com)

Jay Bhatia\*

*Student, Computer Engineering DJ Sanghvi College of Engineering*

Mumbai, India [jaybhatia73@gmail.com](mailto:jaybhatia73@gmail.com)

Pankaj Sonawane\* *Professor, Computer Engineering DJ Sanghvi College of Engineering*

Mumbai, India [pankaj.sonawan](mailto:pankaj.sonawane@djsce.ac.in)[e@djsce.ac.in](mailto:e@djsce.ac.in)

Nimit Vasavat\*

*Student, Computer Engineering DJ Sanghvi College of Engineering*

Mumbai, India [vnimit1991@gmail.com](mailto:vnimit1991@gmail.com)

***Abstract —* Sepsis can cause overwhelming changes that impair multiple organs, leading to their failure and sometimes even having fatal results. Low blood pressure, difficulty in breathing, fever, mental confusion, and fast heart rate are some of the symptoms. Infections acquired in health care settings also frequently result in Sepsis. These infections affect hundreds of millions of patients globally every year and are one of the most recurrent adverse events during care delivery. Clinical conditions can rapidly deteriorate as healthcare- associated disorders are often impervious to antibiotics. During critical scenarios such as an intensive care unit, Early diagnosis and timely medical treatment of sepsis is crucial to give the patient prompt treatment since there is an increase in the mortality rate as each hour passes away in critical care. This paper proposes a method to predict the inception of sepsis 6 hours in advance using various machine learning and deep learning models and presents a comparative study of the same. The Medical Information Mart for Intensive Care III (MIMIC3) dataset was used to test the traditional machine learning methods such as RandomForest(RF), XGBoost, and also Deep Learning techniques such as neural networks and Autoencoders with XGBoost. The dataset contains an extensive range of parameters that are associated with laboratory, vitals, and demographics of patients which help them classify as sepsis and non-sepsis patients. A data pipeline was successfully created to clean the data, impute missing values and perform various feature engineering techniques. Our best performing model is a Deep Neural Network with an AUC ROC score of 0.888.**

***Keywords — Sepsis, Deep neural networks, Autoencoders, XGBoost, Random Forest, Machine Learning***

1. INTRODUCTION

Sepsis can be defined as chronic critical illness related to severe inflammation, reduced immune and organ damage [1]. Sepsis is one of the major causes of death in the Intensive Care Unit (ICU) which affects more than 18 million all over the world with mortality rates as high as 30%. In 2013, the US health system accounted for $24 billion for the treatment of sepsis [2]. Prior detection and specific treatment has been identified as critical features to ameliorate the conditions of patients with sepsis. It can consequently reduce post-traumatic ramifications. A delay in the detection may increase mortality by approximately 4-8% [3]. Due to these reasons, sepsis has become a global public health issue that requires attention in developing ways for early detection of sepsis.

Currently, doctors identify sepsis by using the patient's vital signs and the symptoms, and individualistic biomarkers. But due to the intricacy of sepsis and various organ dysfunction, the biomarkers may differ for every patient [2]. Clinical Decision Rules (CDRs) are currently used for predictive analysis which lacks generalizability when applied to mass populations. Moreover, it may take years for CDRs to develop which makes it difficult for them to update when new information is accessible [4]. There is a requirement to enhance diagnosis and prediction which are more reliable and convenient than the traditional methods.

With the burgeoning of computerized data, machine learning (ML) models have become popular for predictive analysis [4]. A variety of ML models have been used for diagnosis which includes SVM, XGBoost, KNN, improvement in traditional Deep Forest [3], Deep Learning models [5], LSTMs [1], RNNs[4], and many more.

This paper proposes a method that can predict sepsis 6 hours earlier than the clinical prediction. This paper also provides a comparative study of different ML models which are trained and trained on a publicly available dataset. An ML predictive model allows us to identify relations between different variables and discover patterns that are otherwise inconspicuous in traditional predictive methods.

1. LITERATURE SURVEY

The following papers focused on predicting sepsis using Machine Learning and providing results. Md. Mohaimenul Islam et al [6] performs a literature survey of sepsis papers published between 2000-2018 and concludes that the machine learning approach to predict sepsis had a better performance than the traditional sepsis scoring systems such as SIRS, MEWS, and SOFA. Desautels et al [7] have applied the InSight machine learning classification system to predict sepsis using multiple variables and compared its results with systemic inflammatory response syndrome (SIRS), modified early warning score (MEWS), simplified acute physiology score (SAPS) II, and quick sequential organ failure assessment (qSOFA). Their results show that their machine learning model InSight produced superior performance on detecting sepsis onset in comparison with the other methods. They have even shown that the model also performs well in the case of randomly missing data.

M. Saqib et al [2] have compared the results of the traditional machine learning methods, LSTM and Attentional

2021 2nd Global Conference for Advancement in Technology (GCAT) | 978-1-6654-1836-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/GCAT52182.2021.9587778

‘\*’ indicates equal contribution

**978-1-6654-1836-2/21/$31.00 ©2021 IEEE** 1

LSTM for early Sepsis Prediction, and a comparative study is documented. Various statistical techniques such as the chi- squared test of independence were used to determine the best features which were to be used as input data. Some feature engineering techniques were used to determine how some features are directly and inversely related to the Sepsis Label. The proposed method states that RF was the most successful model and LSTM models trained with attentional mechanisms only achieved a slight but noticeable improvement over LSTM models trained without attentional mechanisms.

S. D. Wickramaratne et al [4] have proposed a GRU Network as a model for Sepsis Detection beforehand. All vital, laboratory and demographic data are used by the proposed model to predict sepsis. Their analysis suggests that GRUs make each recurrent unit capture dependencies of different time scales adaptively. The proposed model uses Bi-Directional GRU because over LSTM, due to a lesser number of parameters the GRU model uses less computational time and converges faster, thus providing a better result than LSTM in some cases. The proposed model used two classifiers, one which included only the vital signs as features and the other included both vital signs and laboratory values, the latter giving better accuracy. Traditional baseline models such as Logistic Regression(LR), Random Forests(RF), and Support Vector Machines(SVM) gave a poor performance than The proposed GRU model with regards to sensitivity and specificity.

Bedoya et al [8] have trained a deep learning model which is a multi-output Gaussian process and Recurrent Neural Network(MGP-RNN). Their training set and internal validation included 42 979 encounters, while their temporal validation set included 39 786 encounters. They made comparisons to Random forest, cox regression, penalized linear regression and the clinical scores used to detect sepsis were SIRS, qSOFA, and NEWS. The C-statistic of their model predicting sepsis within 4 hours of onset was higher than the other models. Their model detected sepsis 5 hours early. The temporal validation assessment continued to show higher performance. L. Tran [9] et al has proposed a deep neural network called AEC-Net which concurrently optimizes an auto-encoder and a fully connected neural network. Another model is also proposed which is an ensemble of AEC-Net, RF and GBDT. Both of these

achieved better performance than the baseline models(RF and XGB) taken by the authors.

M. Fu et al [3] have used the MIMIC-III dataset consisting of a total of 3125 patients for early prediction of sepsis and do a comparative study of different machine learning models. Their improved cascade deep forest model includes two Random Forest and two XGBoost in each layer which performs better than the traditional machine learning models. It makes use of k-fold cross-validation.

S. Sarafrazi et al [12] have provided a comparative study of different models to predict sepsis 12 hours before its diagnosis. Various sequential models such as CNN, LSTM, CNN\_LSTM, and XGBoost model are trained with XGBoost outperforming all the other models including XGBoost- CNN-LSTM. Feature Engineering is also performed which results in a 5% improvement in the XGBoost model. G. Tsang et al [5] have proposed a fresh deep learning model for sepsis detection up to 6 hours in advance. The novel model is a combination of cascade and boosting algorithms that improve upon the imbalance of traditional methodologies and provides a more balanced method. The hinge loss function prevents the model from over-fitting over the increasing parameter-space complexity of boosted cascade methods.

1. SYSTEM METHODOLOGY
2. *System Workflow*

The MIMIC-3 Dataset was used for our proposed method. The dataset was first grouped patient-wise for better analysis and further prediction. Huge datasets obtained from public repositories usually contain redundant data which has to be cleaned efficiently. The missing values were imputed patient-wise. Various methods were analyzed for filling in the missing data and the most efficient method was considered. Various data preprocessing steps were used to undersample the data because of a huge imbalance between the positive and negative values. Feature Selection was performed wherein the most important features were considered as input data for the models. Different models were trained and hyperparameter tuning was performed. A comparative study of their results is provided in our proposed method.

Refer Fig 1 to navigate through the system workflow.

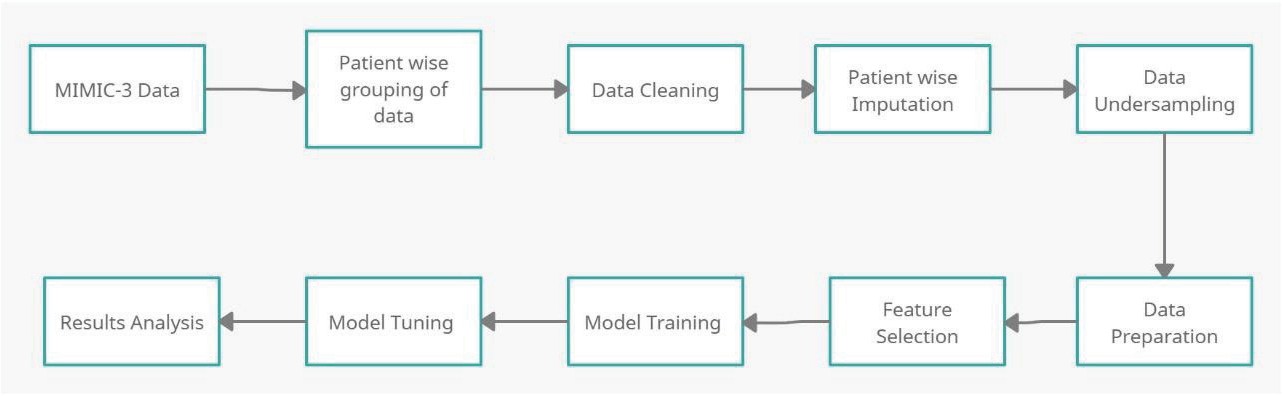


Fig. 1. System Workflow

1. *Dataset Description*

The dataset used is from PhysioNet Computing in Cardiology Challenge 2019 [11] which consists of 40,336 records of patients with 2932 patients detected positive with sepsis and 37,404 patients not detected with sepsis.

This dataset is collected from two different hospitals providing patient-level hourly based records with 40 time- dependent variables divided into three categories - vital signs, lab tests results, and demographic information. The vital signs values were accessible for majority of the records averaging about 32.4% missing data but the majority of the values were missing from the lab results averaging 94.9%. As for the demographic features, there is no data missing.

Target labels are provided for hourly timestamps as either 0 or 1, negative or positive.

The dataset consists of 22,566 males and 17,770 females with an average age of 61.6 years. The minimum and maximum age recorded is 14 and 90+ respectively.

1. *Data Preprocessing*

The MIMIC3 dataset was first extracted and all the individual patient files were combined in comma-separated values (CSV) format. Each row was given a patient id so that the same patient rows can be grouped. The laboratory values were recorded periodically and hence most of them had missing percentages of more than 90%. Fig 2 shows the percentage of missing data in each feature.

Heart rate (HR) and Pulse oximetry (O2Sat) are important features in predicting Sepsis. The patients with more than 50% null values for these two features were removed. Patients with overall missing data greater than 65% were also removed. Since our goal was to predict sepsis for adults, patients with age less than 18 were also removed.

Initially, the remaining missing data were replaced with mean and median for each patient but the results came out

better when the missing data was imputed with forwarding fill and backward fill techniques for each patient.

Forward fill propagates the last observed non-value forward until another non-null value is encountered. Backward fill propagates the first observed non-null value backward until another non-null value is met. Forward fill was applied to missing values in tail rows of a patient and backward fill was applied to missing values in head rows of a patient. After all the above steps, the columns which still had a majority of their values missing throughout the dataset were removed. Four such columns were found namely, End- tidal CO2 (EtCO2), Fibrinogen, Bilirubin direct, and Troponin I (TroponinI). Physiologically normal values were considered for the rest of the features whose entire column was missing for a patient.

Only 24630 patient records were left after preprocessing. But this data is still unbalanced. Among the 24630 patients left, 21698 were negative for sepsis and 2932 were positive for sepsis. If this dataset is trained, our model would be biased. Hence, an undersampling technique was applied to balance the dataset. The sepsis negative patients were randomly deleted so that they can match with the sepsis positive patients. The final dataset contained 5928 patients, out of which 2998 were negative for sepsis and 2930 were positive for sepsis.

Instead of splitting the data by 80-20 split, patient-wise splitting was done to prevent data leakage. In an 80-20 split, the rows of the dataset are split randomly and hence some patients might be present in training as well as testing set. This causes the model to be overfitted. Hence, a patient- wise split is better where all the rows of a particular patient are either present in the training set or the testing set. 5000 patients were used for training and the remaining 928 patients were used for testing.

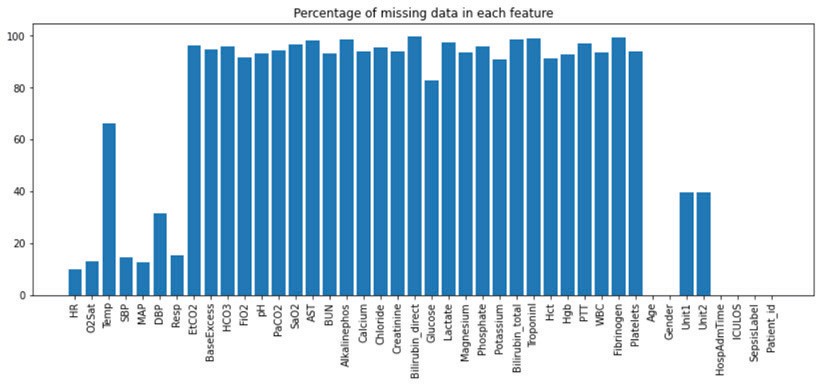


Fig. 2. Percentage of missing data in each column

1. MODEL BUILDING
2. *Traditional Machine Learning Techniques*

Traditional machine learning models like XGBoost and Random Forest were trained. The test set was excluded in the calculation of F scores in feature importance to prevent data leakage. Top 37 features were included in training both the XGBoost and Random Forest models. Table I includes the F scores of the top 14 features. Hyperparameter tuning was done in both models.

For XGBoost, the hyperparameters mostly included booster parameters such as maximum depth, minimum child weight, gamma, subsample, colsample by a tree, and regularization parameters. Learning parameters i.e learning rate was also tuned. Even though the number of positive and negative patients of sepsis were almost equal, the hourly cases for sepsis labels of all patients were still unbalanced. Hence, the booster parameter scale\_pos\_weight was also used to improve results. The value of this hyperparameter was kept as the ratio of negative sepsis labels to positive sepsis labels.

TABLE I. F SCORE OF TOP 14 FEATURES

|  |  |  |  |
| --- | --- | --- | --- |
| **Features** | **F score** | **Features** | **F score** |
| ICU length of stay (ICULOS) | 88 | Creatinine | 25 |
| White Blood Cells (WBC) | 51 | Aspartate transaminase (AST) | 24 |
| Temperature (Temp) | 47 | Oxygen saturation from arterial blood (SaO2) | 24 |
| Hours between hospital admit and ICU admit (HospAdmTime) | 42 | Partial thromboplastin time (PTT) | 23 |
| Platelets | 29 | Blood urea nitrogen (BUN) | 22 |
| Calcium | 28 | Alkaline phosphatase (Alkalinephos) | 17 |
| Fraction of inspired oxygen (FiO2) | 28 | Bilirubin direct | 16 |

For Random Forest, hyperparameters that were tuned included n estimators, minimum samples leaf, split criteria, bootstrap, the maximum number of features kept, maximum depth, and n jobs. Similar to scale\_pos\_weight, class weight hyperparameter was used in Random Forest to tackle the sepsis label imbalance. To further tackle the sepsis label imbalance, a Balanced Random Forest Classifier was used which randomly undersamples each bootstrap sample to

balance it. Better results were achieved by using the Balanced Random Forest Classifier.

1. *Autoencoders with XGBoost*

Autoencoders can be classified as a neural network method that is trained to study a compressed representation of raw data. In the proposed model autoencoder is used as a classification predictive model for the Sepsis Label wherein, the compressed representation of the input features is learned by the model. The design of the autoencoder is such that the input data is restricted to the bottleneck up to the midpoint of the model and then the data is reconstructed generating pseudo input data.

It is an unsupervised approach for learning a lower- dimensional feature representation from training data. Autoencoder consists of encoder and decoder layers. Encoder learns the mapping from data to a low-dimensional latent space where data can be compressed into a small latent vector and compact and rich feature representation can be learned. Decoder learns to map back from latent space to reconstructed training data. Before training the Autoencoder the training data is normalized using a MinMaxScaler. An autoencoder model has been created in which only the non-Sepsis cases. The model will try to learn the best representation of non-Sepsis cases. This is because the autoencoder will try to learn only one class and automatically distinguish the other class.

The reconstruction loss should be as minimal as possible as it forces the latent representation to encode the majority of information about the data as possible into a lower- dimensional latent space while still being able to generate correct representation. Another model consisting of sequential layers is constructed and the trained weights till the layer of the latent representation of the input trained model in the architecture are added in this model. The hidden representation of the sepsis and non-sepsis labels are thus generated by predicting the input data using the above sequential model. The new reconstructed input data generated by the autoencoder is used as training data for the proposed XGBoost Model defined previously and the results of the two models are compared.

1. *Deep Neural Network*

In this section, we present our proposed Deep Neural Network(DNN) model. DNN is a good choice not only due to its property of approximating functions but also due to its feature learning capacity [9]. It is a powerful ML method with a major drawback, overfitting, which we have controlled using early stopping. The model contains three hidden layers with 128, 128, 128 cells respectively, and rectified linear activation or Relu is used to transform summed weighted input from a hidden layer to input to another layer. The Relu function and its derivative both are monotonic. It returns 0 if the input is less than zero i.e. negative and returns 1 if the input is equal to or greater than zero i.e. positive. Hence it ranges from 0 to infinity.

Adam optimizer is used as an optimization technique for gradient descent. The method is effective when the data is large and involves a lot of parameters. It uses an average of second moments of gradients instead of only adapting the learning parameter rates based on the average first-rate.





Eq. (1)

In Eq. (1) w represents weight at a given time, m is the aggregate of gradients at a given time, β is the moving average parameter, is the learning parameter and ߜL represents derivative of a loss function.

1. RESULTS AND DISCUSSION

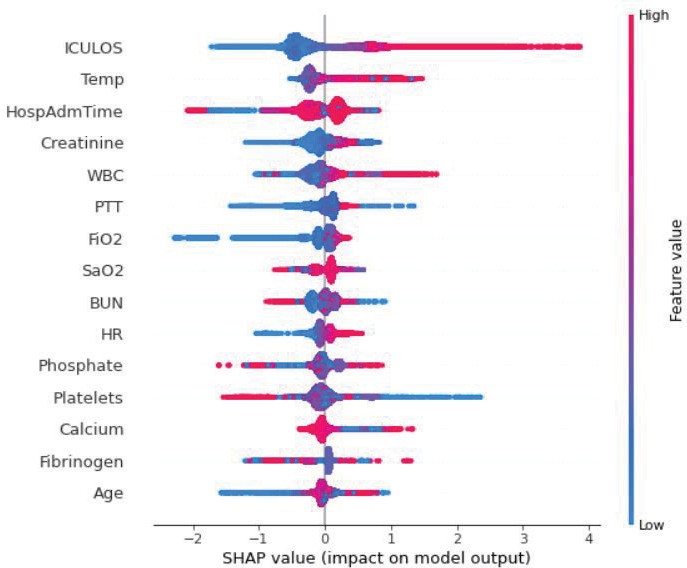
The feature importance was calculated with the help of SHAP values from the SHAP library. The SHAP values can collectively describe how each feature is correlated, either positively or negatively to the target variable. It shows the positive or negative relation of each feature with the target variable. The global interpretability of the features is displayed in Fig 3 and Fig 4 using SHAP values.

Fig. 4 shows the correlation effect of the top 15 features and if they are positively or negatively correlated. The features having a blue bar are positively correlated with the target variable and the features having a red bar are negatively correlated with the target variable. It is observed in Fig 3 that ICULOS has a high positive correlation with the target variable, whereas Phosphate has a high negative correlation with the target variable. It can also be inferred that features such as Temp, HospAdmTime, Creatinine, WBC, etc. have a positive correlation with the sepsis label of varying degrees. Other features such as Platelets, Calcium, and AST have a high negative correlation with sepsis labels.

The four models were evaluated based on five metrics - Accuracy, Sensitivity, Specificity, AUC ROC score, and F1 score.

As seen from Table II., the specificity for traditional machine learning techniques is good but the sensitivity is low and hence, these models cannot be completely trusted for positive label predictions. During the hyperparameter tuning of traditional machine learning techniques, it was observed that there was always a trade-off between accuracy and sensitivity. If the accuracy was increased, the sensitivity decreased, and if the sensitivity was increased then accuracy as well as specificity decreased. The model results shown were the best that could be tuned with a balance of sensitivity and specificity.

TABLE II. RESULTS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | **Sensi- tivity** | **Speci- ficity** | **AUC ROC** | **F1** |
| **XGB** | 74.58 | 0.55 | 0.77 | 0.713 | 0.302 |
| **RF** | 76.12 | 0.51 | 0.79 | 0.712 | 0.301 |
| **Auto-XGB** | 84.31 | 0.58 | 0.87 | 0.816 | 0.417 |
| **DNN** | 81.08 | 0.86 | 0.77 | 0.888 | 0.819 |

Fig. 3. SHAP Value of top 15 Features

In Fig 3 the variables are ranked in decreasing order of their feature importance. The horizontal position for each value in the features displays the association of the value with a higher or lower prediction and the color represents if the variable is high or low for the particular observation.

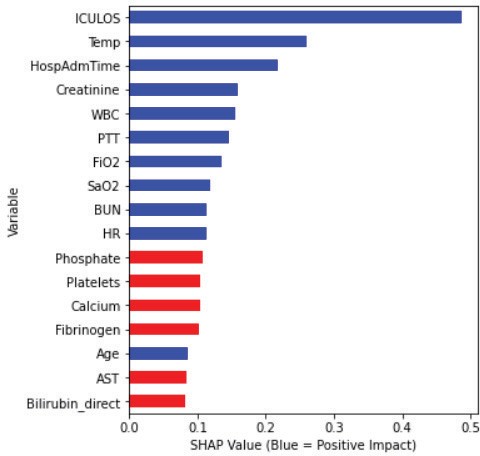


Fig. 4. Correlation Impact of Top Features

By adding Autoencoders to the XGBoost model, the specificity of the model increases but there is not much effect on sensitivity. The accuracy of this model comes out to be the best among all four models.

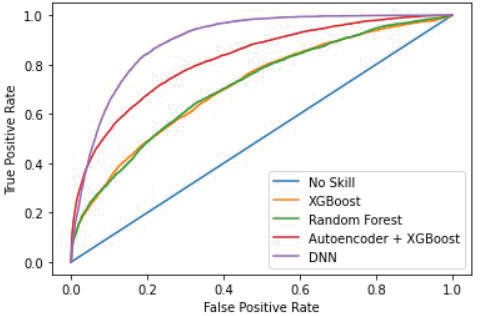


Fig. 5. AUCROC Curve

The Deep Neural Network model built has the highest sensitivity among all the models. Even though the accuracy and specificity of Autoencoder with XGB is higher than Deep Neural Network but the predicted positive label for

sepsis by the Deep Neural Network model will be more reliable due to its high sensitivity than other models. The AUC ROC score of the proposed models is displayed in a graphical format in Fig. 5.

In conclusion, Deep Neural Network is the best performing model out of all the proposed models.

1. CONCLUSION

In our proposed system, the data preprocessing steps such as filling in missing data and filtering data were successfully carried out according to the model requirements, and clean pre-processed data was given as an input to our proposed system to predict sepsis. A Deep Neural Network (DNN) was proposed that beats the traditional machine learning techniques in terms of AUC ROC and F1 score. Autoencoder with XGBoost was also proposed but the DNN outperforms them too. Thus, the proposed system helps clinicians to predict sepsis 6 hours early.

In future work, more sepsis-positive patients can be added to the dataset for better accuracy and more precise models. Using hybrid models with advanced hyperparameter tuning can result in better accuracy of sepsis label prediction and can be more reliable.

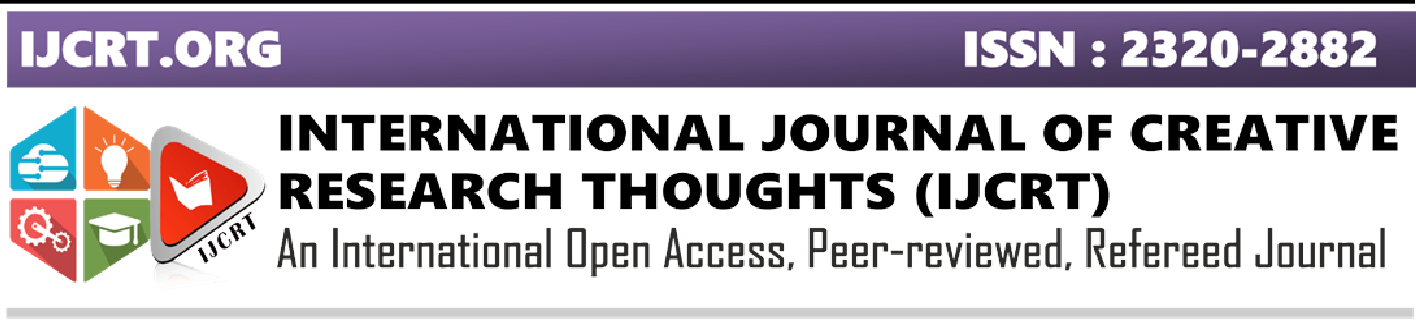
REFERENCES

1. T. Vicar, P. Novotna, J. Hejc, M. Ronzhina and R. Smisek, "Sepsis Detection in Sparse Clinical Data Using Long Short-Term Memory Network with Dice Loss," 2019 Computing in Cardiology (CinC), 2019, pp. Page 1-Page 4, doi: 10.23919/CinC49843.2019.9005786.
2. M. Saqib, Y. Sha and M. D. Wang, "Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 4038-4041, doi: 10.1109/EMBC.2018.8513254.
3. M. Fu, J. Yuan and C. Bei, "Early Sepsis Prediction in ICU Trauma Patients with Using An Improved Cascade Deep Forest Model," 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), 2019, pp. 634-637, doi: 10.1109/ICSESS47205.2019.9040774.
4. S. D. Wickramaratne and M. Shaad Mahmud, "Bi-Directional Gated Recurrent Unit Based Ensemble Model for the Early Detection of Sepsis," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 70- 73, doi: 10.1109/EMBC44109.2020.9175223.
5. G. Tsang and X. Xie, "Deep Learning Based Sepsis Intervention: The Modelling and Prediction of Severe Sepsis Onset," 2020 25th

International Conference on Pattern Recognition (ICPR), 2021, pp. 8671-8678, doi: 10.1109/ICPR48806.2021.9412058.

1. Md. Mohaimenul Islam, Tahmina Nasrin, Bruno Andreas Walther, Chieh-Chen Wu, Hsuan-Chia Yang, Yu-Chuan Li, Prediction of sepsis patients using machine learning approach: A meta-analysis, Computer Methods and Programs in Biomedicine, Volume 170, 2019, Pages 1-9, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2018.12.027.
2. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman M, Barton C, Wales D, Das R Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach JMIR Med Inform 2016;4(3):e28
3. URL: https://medinform.jmir.org/2016/3/e28 [9] DOI: 10.2196/medinform.5909
4. Bedoya, Armando & Futoma, Joseph & Clement, Meredith & Corey, Kristin & Brajer, Nathan & Lin, Anthony & Simons, Morgan & Gao, Michael & Nichols, Marshall & Balu, Suresh & Heller, Katherine & Sendak, Mark & O’Brien, Cara. (2020). Machine learning for early detection of sepsis: an internal and temporal validation study. Journal of the American Medical Informatics Association Open. 3. 10.1093/jamiaopen/ooaa006.
5. L. Tran, M. Nguyen and C. Shahabi, "Representation Learning for Early Sepsis Prediction," 2019 Computing in Cardiology (CinC), 2019, pp. 1-4, doi: 10.23919/CinC49843.2019.9005565.
6. S. Sarafrazi et al., "Cracking the “Sepsis” Code: Assessing Time Series Nature of EHR Data, and Using Deep Learning for Early Sepsis Prediction," 2019 Computing in Cardiology (CinC), 2019, pp. Page 1-Page 4, doi: 10.23919/CinC49843.2019.9005940.
7. Reyna, M., Josef, C., Jeter, R., Shashikumar, S., Moody, B., Westover, M. B., Sharma, A., Nemati, S., & Clifford, G. (2019). Early Prediction of Sepsis from Clinical Data -- the PhysioNet Computing in Cardiology Challenge 2019 (version 1.0.0). PhysioNet.
8. M. Shahul and K. P. Pushpalatha, "Machine Learning Based Analysis of Sepsis: Review," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-4, doi: 10.1109/ic-ETITE47903.2020.399.
9. M. J. Pettinati, G. Chen, K. S. Rajput and N. Selvaraj, "Practical Machine Learning-Based Sepsis Prediction," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 4986-4991, doi: 10.1109/EMBC44109.2020.9176323.
10. A. Shankar, M. Diwan, S. Singh, H. Nahrpurawala and T. Bhowmick, "Early Prediction of Sepsis using Machine Learning," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 837-842, doi: 10.1109/Confluence51648.2021.9377090.
11. F. Mahmud, N. S. Pathan and M. Quamruzzaman, "Early detection of Sepsis in critical patients using Random Forest Classifier," 2020 IEEE Region 10 Symposium (TENSYMP), 2020, pp. 130-133, doi: 10.1109/TENSYMP50017.2020.9231011

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882**



**Accurate Prediction Of Sepsis In ICU Patients.**

Guide - Prof. M. A. Rane

Team - Sneha Bamane, Shweta Maharanawar, Devyani Pathrikar IT Department, BVCOEW, Pune, India

##### ABSTRACT:

Sepsis is a potentially life-threatening condition that occurs when the body's response to an infection causes inflammation throughout the body. This inflammation can trigger a cascade of changes that can damage multiple organ systems, leading to organ failure and death if not treated promptly. Early recognition and aggressive treatment with antibiotics and supportive care are crucial for improving outcomes in septic patients. The " Accurate Prediction of Sepsis in ICU Patients" is a project that combines awareness and predictive modeling to address sepsis, a life-threatening condition commonly encountered in intensive care units (ICUs). This project is a robust awareness campaign designed to educate both the general public and healthcare professionals about sepsis. With focusing on generating awareness about sepsis, can lead to early detection and seeking medical help. By bringing the limelight on this disease, it can potentially save lives. Concurrently, advanced machine learning techniques, specifically random forest algorithms, are employed to construct a predictive model for sepsis. This model undergoes meticulous fine-tuning to ensure accurate identification of sepsis risk in ICU patients. It uses a dataset for training the predictive model. The integration of the Sequential Organ Failure Assessment (SOFA) score, including the quick SOFA (qSOFA) criteria, enhances predictive accuracy. The qSOFA criteria play a crucial role in rapid risk assessment for early intervention. Moreover, the project maintains a dedicated website that serves as an essential platform for sepsis education and the dissemination of the predictive model to the medical community.

Keywords: Sepsis, ICU, SOFA, qSOFA, Random Forest, Predictive Modelling, Awareness Campaign

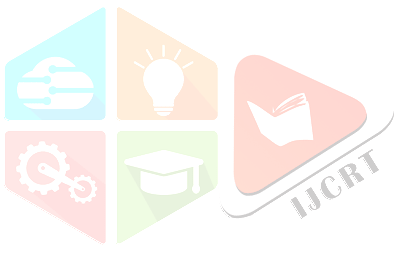
##### INTRODUCTION

Sepsis, a life-threatening condition triggered by infections, remains a critical concern within the realm of intensive care units (ICUs). It is a serious medical condition that can occur due to bacterial, viral, or fungal infections in the bloodstream. When a person has an infection, it causes inflammation in the body, which releases a chemical called cytokines. These cytokines have different effects on the body's tissues, including recruiting immune cells to fight the infection. However, some cytokines can also make blood vessels leaky, causing white blood cells to move from the bloodstream to the tissues. This process leads to the release of heat, redness, and swelling in the affected area. If sepsis

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882** is left untreated, it can progress to severe sepsis and septic shock, which can be fatal. In response to this issue, we present the " Accurate Prediction of Sepsis in ICU Patients" project. This initiative is driven by a dual commitment: raising awareness about sepsis and harnessing the power of predictive modeling to revolutionize patient care within ICUs.

At its core, it prioritizes education and awareness. Our efforts extend to both the general public and healthcare professionals, as we recognize the significance of knowledge in the battle against sepsis. By raising awareness about the gravity of sepsis and the paramount importance of early detection and intervention, we aim to empower individuals to recognize its signs and seek immediate medical attention.

In parallel, the project leverages advanced machine learning techniques, specifically logistic regression algorithms, to construct a predictive model for sepsis. The model's rigorous finetuning is aimed at achieving accurate identification of sepsis risk in ICU patients. This predictive tool represents a critical advancement in healthcare, offering the potential for early interventions and improved patient outcomes.

Our data-driven approach encompasses comprehensive data pre-processing, addressing class imbalances within the dataset. Additionally, the integration of the Sequential Organ Failure Assessment (SOFA) score, including the quick SOFA (qSOFA) criteria, enhances predictive accuracy by evaluating organ failure trajectories and risk factors.

Moreover, the project maintains a dedicated website, serving as an essential platform for sepsis education and the dissemination of the predictive model to the medical community.

##### LITERATURE REVIEW

Anurag Shankar, Mufaddal Diwan, Snigdha Singh, Husain Nahrpurawala and Tanusri Bhowmick [1] underscore the significance of the Sequential Organ Failure Assessment (SOFA) score and elucidate the challenges inherent in achieving timely sepsis diagnosis. This seminal work sets the stage for subsequent research endeavours by highlighting the potential of ML in addressing these challenges.

B. C. Srimedha, Rashmi Naveen Raj, And Veena Mayya [2] present a thorough investigation into four prediction algorithms—Random Forest, Logistic Regression, Gradient Boosting, and Decision Tree—while also examining the impact of various imputation techniques. This comprehensive pipeline offers valuable insights into the development of robust ML models for accurate early prediction of sepsis in ICU settings.

Guilan Kong, Ke Lin, and Yonghua Hu [3] focus on leveraging ML techniques to predict in-hospital mortality among sepsis patients in the ICU. The study develops ML models, including the Least Absolute Shrinkage and Selection

Operator (LASSO), Random Forest (RF), Gradient Boosting Machine (GBM), and Logistic Regression (LR), demonstrating their effectiveness in mortality prediction.

Yash Veer Singh, Pushpendra Singh, Shadab Khan, and Ram Sewak Singh [4] propose a novel ML model for early sepsis prediction in ICU patients, leveraging data from clinical laboratory values and vital signs. By comparing various models, such as Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), and XGBoost, the study identifies an ensemble method that shows promising results in terms of classification performance and prognosis improvement.

Longxiang Su, Zheng Xu, Fengxiang Chang, Yingying Ma [4] focus on harnessing ML techniques to anticipate inhospital mortality among sepsis patients in the ICU. The study develops several ML models, including LASSO, RF,

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882** GBM, and LR, to predict mortality, severity, and length of stay, contributing to a comprehensive understanding of sepsis prognosis.

##### SYSTEM ARCHITECTURE

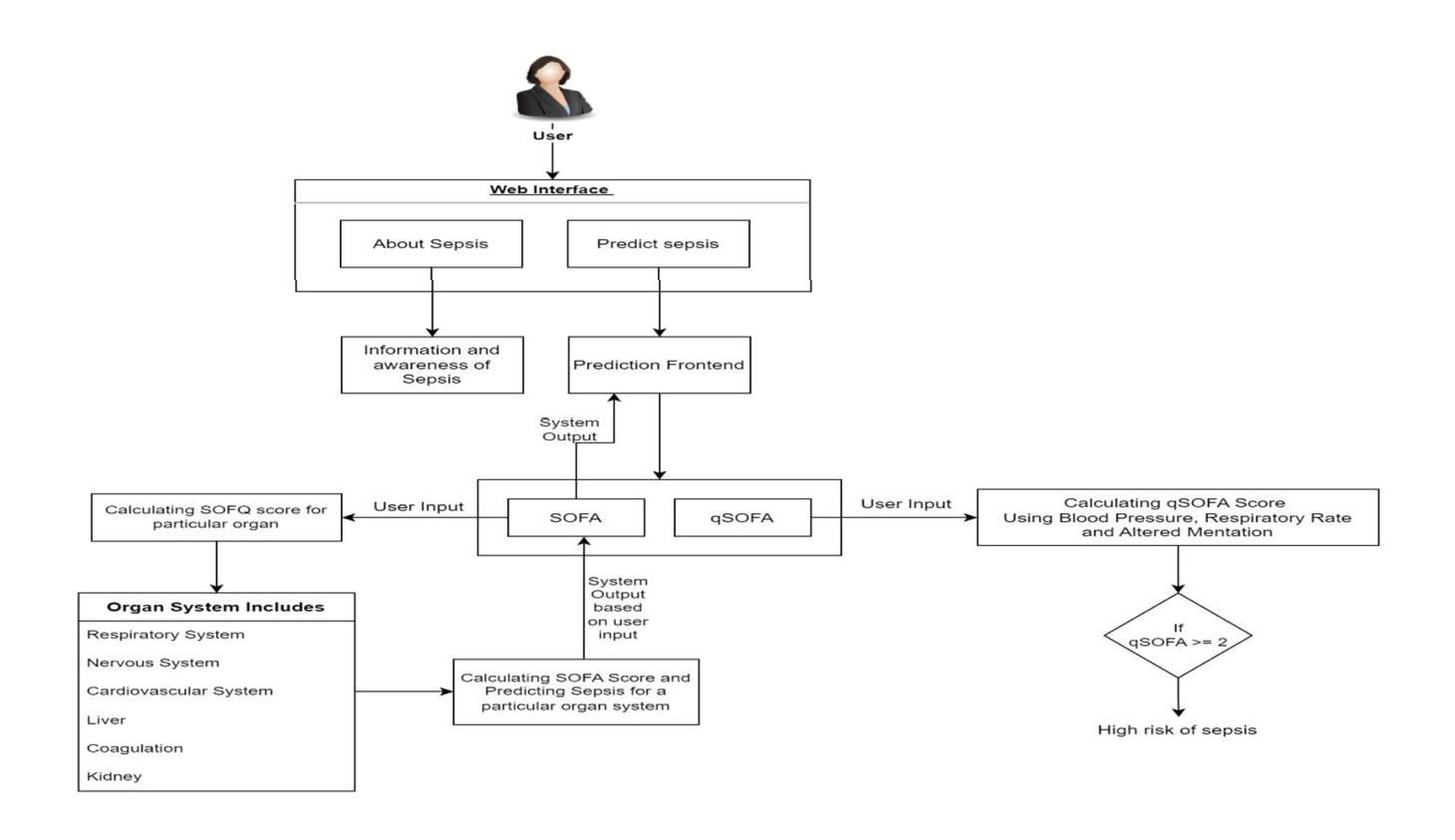
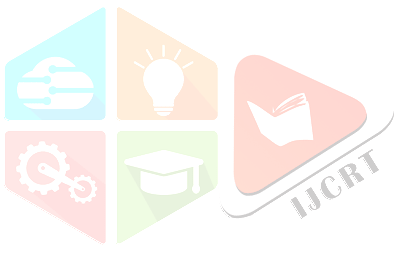


Fig 1. System Architecture

The system architecture for the Accurate Prediction of Sepsis in ICU Patients is designed to seamlessly integrate technology, education, and clinical insights to address the pressing issue of sepsis in intensive care units (ICUs).

1. Input Data: The architecture begins with the input data, which comprises a set of vital signs and clinical parameters from ICU patients like Blood Pressure, Respiratory Rate, Heart Rate, etc.
2. SOFA Score Calculation: The SOFA score is a comprehensive assessment that evaluates the patient's condition across various organ systems, which include Respiratory System, Nervous System, Cardiovascular System, Liver, Coagulation, and Kidney. Machine learning algorithms are employed to calculate the SOFA score by considering the values of these organ systems. The resulting score indicates the patient's overall health.

Machine learning algorithms are employed to calculate the SOFA score based on the values of these organ systems. These algorithms process the input data to assess the severity of organ dysfunction, providing an overall indication of the patient's health status.

1. qSOFA Score Calculation: In contrast to SOFA, the qSOFA score is a more rapid assessment that focuses on three key criteria that are Blood Pressure, Respiratory Rate, and Altered Mentation. We use machine learning techniques to calculate the qSOFA score, enabling quick predictions of sepsis risk. 4. Prediction Outcome: Both the SOFA and qSOFA scores play a crucial role in our sepsis prediction. These scores are used as features to train machine learning models. The architecture integrates these scores with other clinical insights to enhance prediction accuracy.
2. Prediction Outcome: Both the SOFA and qSOFA scores play a crucial role in predicting sepsis risk in ICU patients. These scores, along with other clinical insights derived from the input data, serve as features for training machine learning models.

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882**

Machine learning models are trained using datasets that incorporate patient data along with corresponding sepsis outcomes. These models learn patterns and relationships within the data to predict the likelihood of sepsis development in ICU patients accurately.

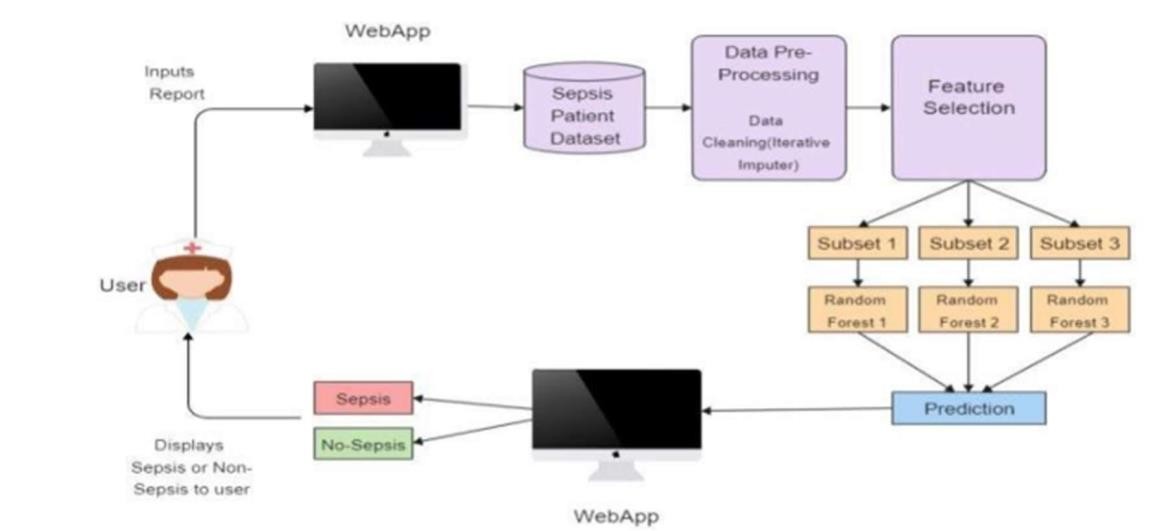
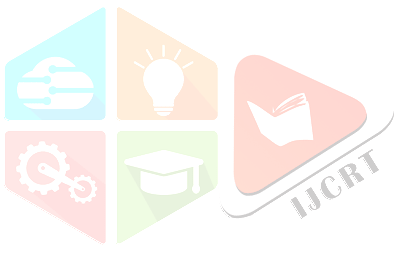
The integration of SOFA and qSOFA scores, along with other clinical features, enhances the prediction accuracy of the models, enabling early detection and intervention to mitigate the risks associated with sepsis.

Machine learning techniques are utilized to calculate the qSOFA score, enabling rapid assessment and prediction of sepsis risk. This score provides a quick indication of the likelihood of sepsis in ICU patients.

##### 4. METHODOLOGY

Web Interface:

The web interface is designed to be intuitive and user-friendly, accessible to both healthcare professionals and the general public. It features a predictive model powered by the Random Forest algorithm, trained on a comprehensive dataset of ICU patient data encompassing vital signs, laboratory results, and demographic information. Additionally, the interface integrates educational content on sepsis, including information on its awareness, symptoms, underlying causes, and precautionary measures.



Functionality:

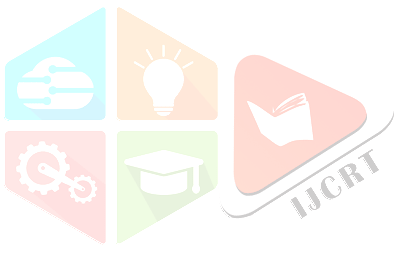
Upon accessing the web interface, users are greeted with a dashboard providing options for sepsis severity prediction and educational resources. The prediction module allows clinicians to input patient data and receive predictions on sepsis severity, categorized as low, medium, or high risk. The results are presented in a clear and interpretable format, aiding clinicians in making informed decisions regarding patient care. Simultaneously, the educational section of the

interface offers comprehensive information on sepsis, covering its definition, epidemiology, risk factors, common symptoms, underlying causes, and recommended precautions. Interactive features such as infographics, videos, and quizzes enhance user engagement and facilitate a better understanding of sepsis-related concepts.

Fig 2. Web Layout

1. Educational Content: The website serves as a platform for educating both the general public and healthcare professionals about sepsis. This could include articles, infographics, videos, and other resources to raise awareness about the condition and its early detection.

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882**

1. Explanation of Predictive Model: The website should provide an overview of the predictive model employed in the project. This explanation could include details on the use of logistic regression algorithms, the integration of SOFA and qSOFA scores, and how these factors contribute to accurate prediction of sepsis risk.
2. Data Input Interface: Users may interact with the website to input relevant data for predicting sepsis risk. This interface could include fields for entering vital signs and clinical parameters, such as blood pressure, respiratory rate, heart rate, etc.
3. Visualization of SOFA and qSOFA Scores: The website can display the calculated SOFA and qSOFA scores for users to better understand the patient's overall health status and sepsis risk.
4. Prediction Outcome Display: Users can view the prediction outcome generated by the machine learning model based on the input data and calculated scores. This could include a risk assessment of sepsis and recommendations for further action.
5. Accessibility Features: Ensure the website is accessible to a wide audience, including individuals with disabilities. This may involve implementing features such as alternative text for images, keyboard navigation, and compatibility with screen readers.
6. Flask Web Framework: The Flask web framework serves as the backbone of the digital platform. It handles user requests, routes them to the appropriate components, and ensures seamless communication between the user interface and the server.
   * Feature selection:

Feature selection helps to select some specific features from the set of features which helps the model to make predictions.

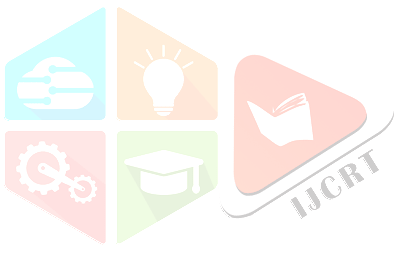
|  |  |  |
| --- | --- | --- |
| Sr. No | Feature | Description |
| 1. | HR | Heart Rate(Beats Per Minute) |
| 2. | DBP | Low (Diastolic)Blood Pressure(mm Hg) |
| 3. | SBP | High (Systolic)Blood Pressure(mm Hg) |
| 4. | O2Sat | Pulse oximetry (%) |
| 5. | Altered Mentation | Glasgow Coma Scale(GCS) |
| 6. | RR | Respiratory Rate (Breaths Per Minute) |
| 7. | Temp | Temperature (Deg C) |

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882**

|  |  |  |
| --- | --- | --- |
| 8. | Liver | Bilirubin(mg/dL) |
| 9. | Creatinine | Kidney Function(mg/dL) |
| 10. | Coagulation | Platelets (count \* 10\*3 / μL ) |
| 11. | Respiratory System | Respiratory System(PaO2/FiO2) |

These parameters encompass physiological, clinical, and laboratory measurements that are relevant for assessing sepsis severity and predicting patient outcomes. Feature selection involves identifying the most informative subset of these parameters to build predictive models for sepsis prediction.

* + Model Training:

1. Dataset Splitting:
   * The dataset containing ICU patient data, including features such as vital signs, laboratory results, and demographic information, is divided into two subsets: a training set and a testing set.
   * The training set is used to train the Random Forest, while the testing set is kept separate for evaluating the trained model's performance.
2. Training the Random Forest Model:
   * With the training set prepared, a Random Forest is trained using this data.
   * Random Forest builds an ensemble of decision trees, where each tree is constructed using a bootstrapped sample of the training data (sampling with replacement).
   * At each node of each decision tree, a random subset of features is considered for splitting, rather than using all available features. This randomness helps to de-correlate the trees and reduce overfitting.
3. Hyperparameter Optimization:

* The Random Forest classifier has several Hyperparameter that control its behaviour during training. Key Hyperparameter include:
* Number of trees (n\_estimators): Determines the number of decision trees in the forest. Increasing the number of

trees may improve model performance but also increase computational cost.

* Maximum depth of trees (max\_depth): Specifies the maximum depth allowed for each decision tree. Deeper trees can capture more complex relationships in the data but may lead to overfitting.
* Minimum samples per leaf (min\_samples\_leaf): Sets the minimum number of samples required to be at a leaf node. This parameter helps control the size of the leaves and prevent overfitting.
* These Hyperparameter can be optimized through techniques like grid search or random search, where different combinations of Hyperparameter values are tested, and the best combination is selected based on performance metrics.

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882**

##### RESULTS

In our study on the accurate prediction of sepsis in ICU patients, we compared the performance of four different prediction models: Random Forest (RF), Naive Bayes, Logistic Regression, and Support Vector Machine (SVM). Each of these models was trained using a dataset containing ICU patient data, including vital signs, laboratory results, and demographic information. Approximate accuracy of models is mentioned below:

* + Random Forest (RF): 75%
  + Naive Bayes: 68%
  + Logistic Regression: 64%
  + Support Vector Machine (SVM): 63%

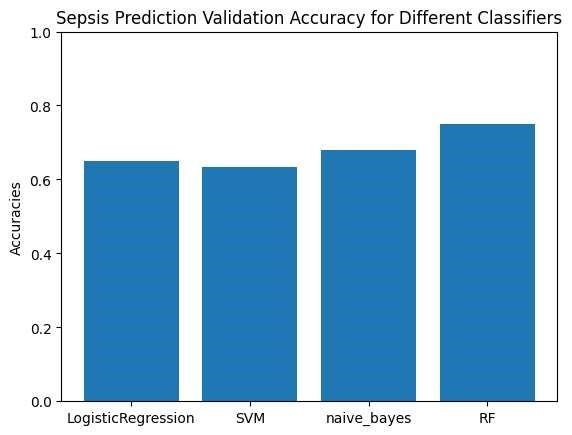
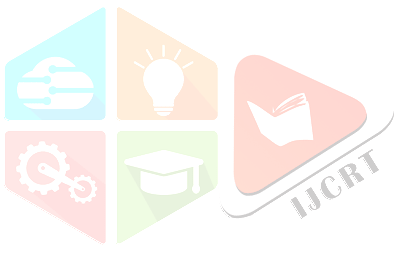


Fig 3. Accuracy Comparison

##### CONCLUSION

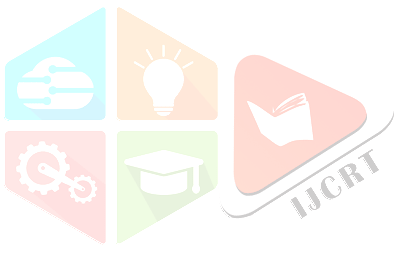
In conclusion, our project represents a pioneering endeavour that addresses the critical issues of sepsis awareness and early detection in intensive care units (ICUs). By integrating education, technology, and clinical insights, this project strives to create a holistic approach to sepsis management, with the ultimate goal of improving patient outcomes and reducing sepsis-related fatalities.

Through a user-friendly web interface and educational content, the project empowers individuals, both healthcare professionals and the general public, with the knowledge and tools to recognize sepsis symptoms early. This educational component, combined with the implementation of the Random Forest Regress or algorithm fine-tuned for accuracy, ensures rapid and accurate sepsis risk assessment.

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882** The project's significance is underscored by the potential to save lives and contribute to a healthier future. By raising sepsis awareness, fostering early detection, and ensuring the ethical and secure handling of healthcare data, the project sets a precedent for the successful integration of education and technology in healthcare practices.

As we look ahead, the project not only holds the promise of reducing the burden of sepsis but also serves as a model for addressing critical medical conditions through interdisciplinary collaboration. With its commitment to improving patient-centered care, the project stands as a testament to the potential of innovation in healthcare, setting a course toward a brighter and healthier future for all.

##### FUTURE SCOPE

In the future, advancements in sepsis prediction and early intervention hold significant promise for improving patient outcomes and reducing the burden on healthcare systems. The integration of real-time monitoring, predictive analytics, and wearable health devices offers opportunities for continuous monitoring and early detection of sepsis outside the ICU setting, enabling timely interventions and potentially preventing the progression to severe sepsis or septic shock. Personalized risk stratification models, informed by genomic and proteomic analysis, can enhance the accuracy and specificity of sepsis prediction, allowing for targeted interventions and tailored treatment approaches. Moreover, the integration of telemedicine platforms facilitates remote monitoring and consultation, extending the reach of sepsis prediction systems to diverse patient populations and geographic regions. Collaborative efforts between clinicians, data scientists, and technology developers will be essential to drive innovation in sepsis prediction algorithms, refine predictive models, and ensure seamless integration with electronic health records for efficient clinical decision support. Patient engagement and education initiatives will play a crucial role in raising awareness about sepsis prevention and management, empowering individuals to recognize early warning signs and seek timely medical assistance.

##### REFERENCES

1. Anurag Shankar, Mufaddal Diwan, Snigdha Singh, Husain Nahrpurawala and Tanusri Bhowmick, “Early Prediction of Sepsis using Machine Learning,” IEEE 2021.
2. B. C. Srimedha, Rashmi Naveen Raj, And Veena Mayya, “A Comprehensive Machine Learning Based Pipeline for an Accurate Early Prediction of Sepsis in ICU”, IEEE access 2022.
3. Guilan Kong, Ke Lin, and Yonghua Hu,” Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU”, 2020, Published by BMC.
4. Yash Veer Singh, Pushpendra Singh, Shadab Khan, and Ram Sewak Singh, “A Machine Learning Model for Early Prediction and Detection of Sepsis in Intensive Care Unit Patients”, 2022, Published by Hindawi.
5. Longxiang Su, Zheng Xu, Fengxiang Chang, Yingying Ma, “Early Prediction of Mortality, Severity, and Length of Stay in the Intensive Care Unit of Sepsis Patients Based on Sepsis 3.0 by Machine Learning Models”, 2021, Published by Frontiers.

[**www.ijcrt.org**](http://www.ijcrt.org/) **© 2024 IJCRT | Volume 12, Issue 5 May 2024 | ISSN: 2320-2882**

1. S. Kanaga Suba Raja, K. Valarmathi, S. Deepthi Sri, “Sepsis prediction using ensemble random forest”, 2022, Published by AIP Conference Proceedings.
2. S. Babu, Lalitha Anupama Annavarapu, Lakshmi Lahari Appala, “Sepsis Detection using Neural Networks”, 2022, Published by SMART GENCON.

