Web Interface for Sepsis Detection

Yash Puri
Department of Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad
Uttar Pradesh, India
yash.2024cs1194@kiet.edu

Vishal Verma
Department of Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad
Uttar Pradesh, India
vishal.2024cs1173@kiet.edu

Yash Goel
Department of Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad
Uttar Pradesh, India
yash.2024cs1050@kiet.edu

Neha Shukla
Department of Computer Science
KIET Group of Institutions
Delhi-NCR, Ghaziabad
Uttar Pradesh, India
neha.shuka@kiet.edu

***Abstract* —This research paper discusses the use of machine learning techniques and their implementation to predict sepsis at an early stage. The study examines different machine learning algorithms that can be used to predict sepsis at an early stage along with the evaluation based on various parameters such as their performance and accuracy. Sepsis is activated by the immune system present in our body that works all the time to protect our body from various possible infections from entering. The main work of our immune system is to fight against bodily infections that may be due to bacteria or various other reasons and during this stage, enormous amount of synthetic substances are discharged into the blood. Sepsis occurs when body's response to these chemicals is out of balance, which can damage multiple organ substances. For the patient the practicality of predicting sepsis disease occurrence in development is an important factor in the result. The primary goal of this study is to examine models using different machine learning algorithms and choose the best one to detect sepsis disease in minimal time. Our secondary goal is to build and design a user- friendly web application. The research concludes that random forest algorithm provides an efficient and accurate way to predict sepsis at an early stage.**

Keywords -- Sepsis, Sepsis Prediction, Logistic Regression, KNN, Naive Bayes and Random Forest.

1. INTRODUCTION

This research paper explores the development and practicality of using machine learning approach for detecting sepsis to improve the treatment outcomes. The paper aims to provide an efficient way of predicting sepsis in minimal time highlighting its potential benefits and future work. This paper discusses the various benefits of implementing machine learning approach instead of using traditional time consuming approach to help doctors for detecting sepsis at an early stage by improving the patient outcomes. The web interface will help health care providers and medical professionals to detect sepsis at a very early stage and help to proceed with required treatment and actions to reduce the mortality rate and other harmful effects caused from sepsis.

Sepsis is a hazardous medical condition that happens when the body's reaction to any contamination which can be because of any external bacteria or other factors causes tissue harm such as damage to organ functions, organ failure, or even death of the person in extreme circumstances. Generally, our body's immune system releases natural chemicals into the blood circulation system in order to fight with the infection which is present inside. Sepsis is a condition which occurs when the body's response to these chemicals is out of balance, this can damage many organ systems. The main cause of sepsis is any kind of infection and can happen to anyone. Sepsis can affect anyone regardless of their age and other factors but it is most commonly found mostly in post op cases because they are more vulnerable to infections also in senior citizens, pregnant ladies, kids below one-year-old, persons suffering from chronic conditions such as diabetes, kidney disease, lung disease, or even cancer, as they have weak immune systems. Nearly 30 percent of patients admitted to ICU in India had sepsis out of which one from 3 died. A study in 2017 also reported that it contributes to substantial deaths in India. This disease is a major health concern for the public in terms of morbidity, health care expenses and mortality. Detecting at early stages is very crucial for improved outcomes of treatment, and providing the patients with antibiotics and other necessary treatment. It can be treated almost with full recovery if it is recognized at an early stage. Several examinations have demonstrated that delays in finding and treatment of sepsis can prompt high death rates or severe after affects.

Overall the main motive of this paper aims to contribute towards the healthcare communities and doctors for improving the probability of detecting sepsis as soon as possible when the patient visits the emergency department in the hospital for treatment.

2. LITERATURE REVIEW

Reference	Major Contributions	Objective	Year	Domain/ Stack	Result
[1]	Predicting Infections Using Computational Intelligence.	To develop computation models using ml algorithm for predicting sepsis by infections and surgical infections.	2020	Machine Learning	The system learns for predicting sepsis mainly by infections and surgical infections.
[2]	A Deep Learning-Based Sepsis Estimation Scheme.	The objective was to design a machine learning based technique that can predict cases of septic shock and extreme sepsis.	2020	Machine Learning	Constructed a model to predict septic and extreme sepsis.
[3]	Trans-thoracic echocardiography and mortality in sepsis.	To examine the role of TTE with 28- day mortality in a population.The MIMIC- 3 database to Identify patients with sepsis who had and had not received TTE.	2018	Machine Learning	Suggested useful diagnostic tool for clinical decision and support.
[4]	Learning Representations for the early detection of sepsis with deep neural networks.	To provide an effective early-stage sepsis detection model using deep learning algorithms.	2017	Machine Learning	Provided an effective system for prediction but complex and may cause different results.
[5]	New Effective Machine Learning Framework for Sepsis Diagnosis.	Provide method which has got 81.6% recognition rate, 89.57% sensitivity and used effectively to diagnose sepsis. sepsis.	2018	Machine learning	It sometimes provides false results resulting in unnecessary treatment.
[6]	Early prediction of sepsis based on ML techniques.	Used XG boost and light gbm to predict sepsis 6 hours in advance.	2021	Machine learning	If the data used is biased it may result in unreliable prediction.

[7]	Multi branching CNN for predicting sepsis.	Multi branching framework to model complex clinical data to predict sepsis.	2021	Machine learning	It becomes complex when the data becomes very large.
[8]	Predicting of sepsis patients using machine learning approach.	Study found that machine learning models performed better than other existing systems.	2019	Machine learning	It doesn't specify that which model is best for predicting sepsis patients.
[9]	Ensemble ML model for early prediction of sepsis.	Proposed an ensemble model by using bagging and boosting trees for predicting sepsis.	2019	Machine learning	It may produce inaccurate predictions.
[10]	Early prediction of sepsis for ICU patients.	Proposed system intended to find and test ml algorithms for prediction.	2022	Machine learning	Implementing to real world may face challenges in Terms of acceptance and integration.

TABLE-1

The comparison of several research papers which are discussed in table 1 are given below:-

In [1], Baldominos et *al*. stated that machine learning techniques such as SVM, Logistic Regression, Naïve Bayes were effective in detecting infections mainly focused towards predicting sepsis. This study mainly focused on how effectively infections can be predicted including sepsis also by using various machine learning algorithms but its implementation cost was expensive because it may involve restructuring of health care systems. In other study use of some specific machine learning model was emphasized to predict cases of sepsis and septic shock by highlighting the relevance and importance of laboratory values of the affected patients but as layers become more the accuracy becomes less which was proposed by Al-Mualemi in [2]. But later on in a similar study an advance and evolved version of C.N.N was proposed that was multi branching C.N.N a novel predictive framework to predict sepsis [7] and also handle the missing values and other data issues but it also had a limitation that the framework becomes complex to operate effectively when the data becomes large. A study focused on developing a deep learning based model for predicting sepsis and also achieved better performance with input to output layer by feeding it in one direction [4], but due to the complexity of the model it was difficult for the model to generate same results for sepsis on different datasets which makes it challenging to compare its performance with the others.

In 2019, a study proposed an idea for use of ensemble machine learning model by using boosting and bagging techniques for predicting sepsis [9]. Both techniques are used for correcting errors , optimizing the models performance , enhancing accuracy. Boosting helps to remove and correct errors sequentially that were made by the previous ones whereas bagging technique works for removing the overfitting problem by averaging the results of different machine learning models but this model was highly dependent on the quality of the input data due to which poor quality of input can ultimately result in inaccurate predictions and poor performance of the model. Similarly a study proposed the use of XG boost and light gbm algorithm to predict sepsis 6 hours in advance [6]. Light gbm is also a gradient boosting framework which uses decision trees to increase the performance also it has fast iteration in training which gives the model a better predictive ability to detect sepsis but if the data provided is biased or we can say the data is incomplete then the accuracy of the model can be affected ultimately leading to unreliable results.

In [3], Feng et *al*. it focused on matching the relationship between trans-thoracic echocardiography and patient within septic addressing the use of TTE but performing TTE can be time consuming which can cause delay in the required necessary treatment or prevention.

In [10], The conducted study was intended to propose a system for predicting sepsis by finding and testing out various potential machine learning algorithms such as decision trees, gradient boosted trees but these models were less accurate and may face some challenges in terms of acceptance and integration with healthcare organizations. In a study feature selection was performed on random forest model before building their classification model to optimize it which was then used to effectively diagnose the sepsis but the proposed model has lower specificity rate that means it is more likely to produce false results which will ultimately lead to unnecessary treatment [5]. In a study it was stated that machine learning models are proved to be superior in identifying and predicting sepsis or other infections also over other existing systems and will have very less false results if implemented correctly which can lead to improve the patient outcomes who are suffering from sepsis but the study did not suggested that which machine learning model would be best for predicting sepsis [8].

These all studies that are discussed above as a part of literature review clearly indicates the effectiveness and future scope of various advancements that can be done in predicting sepsis easily by using various machine learning models and hybrid techniques. Various further studies were made to make the performance of existing machine learning models much better or to introduce new methods. After looking on the immense potential of machine learning techniques for predicting infections in this paper we are trying to find out a superior machine learning model which would be best for predicting sepsis accurately at an early stage.

3. PROPOSED APPROACH

The given approach has been proposed to enable the health care community to predict sepsis at an early stage because it is a one of the leading cause of mortality. This approach can help to predict sepsis at an early stage and reduce the after effects as much as possible which are caused by sepsis. This approach basically compares the performance of different machine learning algorithms and find out the best one that can be used to predict sepsis infection efficiently. This is specifically designed for health care organizations to provide an ease for detection of this hazardous condition and plan accordingly.

This research aims to analyse an effective sepsis detection model using a data set which is obtained from Physionet, comprising clinical data from ICU patients across multiple hospitals. The data-set consists of vital signs, laboratory values, and other important parameters for over 40,000 patients, with a significantly ectreme imbalance between non-sepsis and sepsis cases (97.8% vs. 2.2% respectively). The proposed approach consists of several steps that are given below:-

1) Data Pre-processing:
Preprocess the data by removing null values or missing values in order to overall balance the data-set also by addressing by dropping those values with over 80 percent missing values. Resampling the data set to address the imbalance between sepsis and non-sepsis cases.

2) Algorithm Implementation:
Splitting the preprocessed data set into two parts that is training and testing subsets.Employing various classifiers (Logistic Regression, KNN, Naive Bayes, Random Forest) to train models on the training data set.The roles of these classifiers are given below:-

- Logistic regression = It is a classification model that's incredibly simple to implement and perform admirably well for predicting the probability of something happening or not happening. It is a powerful tool which can be used for binary classification tasks providing certain predictions depending on the situation.

- KNN = KNN, or k-nearest neighbors, is one of the most simplest algorithms and works by looking for making predictions by majority class. It works by looking for similar patients when a new patient comes in the case of sepsis prediction. When a new patient comes in, KNN checks how similar their information is to past patients who either had or didn't have sepsis. It then looks at the closest matches that are found based upon these factors. KNN compares new patient's vital signs, and other parameters to past patients to make a guess about the likelihood of sepsis.

- Naïve Bayes = Naïve Bayes Classifier is one of the simple and most effective classification algorithms which helps in building the fast machine learning models that can make quick predictions. When a new patient arrives, Naive Bayes looks at their vital signs and other values then calculates the likelihood of sepsis based on each clue separately. Then it combines all these individual probabilities to make an overall prediction about whether the patient might have sepsis or not.

- Random Forest = Random Forest is a popular machine learning algorithm which is widely used to improve performance and belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of combining multiple classifiers to solve complex problem and improve performance. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

3) **Performance Evaluation:** Assessing the model performance based on various factors such as accuracy, precision, recall, F1 Score, AUC-ROC, Mean Absolute Error, and Root Mean Square Error. And identifying the best- performing classifier among all the considered models.

4) **Development of User Interface (GUI):** Designing a user-friendly GUI using Flask to allow users to upload text files containing relevant patient vital signs and other features. Implementing a prediction mechanism to determine whether a patient is likely to have sepsis or not based on the uploaded data.

In the proposed approach that we have discussed in this paper have several benefits towards clinical sector for detecting early stage sepsis and for using this the healthcare organization will have to provide the patient data which will have various vital parameters such as age, gender, time under treatment in ICU for any particular condition, respiratory rate, blood laboratory values (platelets, calcium, parameters that are related to other organs working) and predict sepsis efficiently. After providing with the parameters and lab information related with the patient, the data will be processed by the proposed system and will predict whether the patient has sepsis or not. If the patient will have sepsis then the interface will show sepsis label as 1 otherwise 0 or everything looks fine. The data of the patient will be processed by various machine learning algorithms (Naïve Bayes, Random Forest etc) and will predict the sepsis by using that algorithm which will have highest level of accuracy. In this approach we are trying to find out the most accurate algorithm to predict sepsis efficiently. The doctors of health care organization don't need to sign up they just need to provide the data related to the patient for whom they want to detect sepsis. The health care providers can easily view the results for sepsis detection in minimal time with accuracy. There will be lots of benefits for both doctors as well as for patients also because this system can help to predict sepsis with in minimal time without any human intervention also there's a very less possibility of getting false results. It will help doctors to treat sepsis rapidly by taking necessary actions because sepsis is a condition which can be treated with a nearly full recovery and no long-term health issues if it is detected at an early stage. By adopting this approach at a large scale there's a huge possibility to see a significant decrease in mortality rates due to sepsis. These steps mentioned above for the approach aims to systematically process the data set, develop machine learning models, and create an accessible simple user interface for real-time sepsis prediction, contributing towards the betterment of enhanced clinical diagnostics.

4. ARCHITECTURE

This section will include data flow diagram for our proposed system and the architecture.

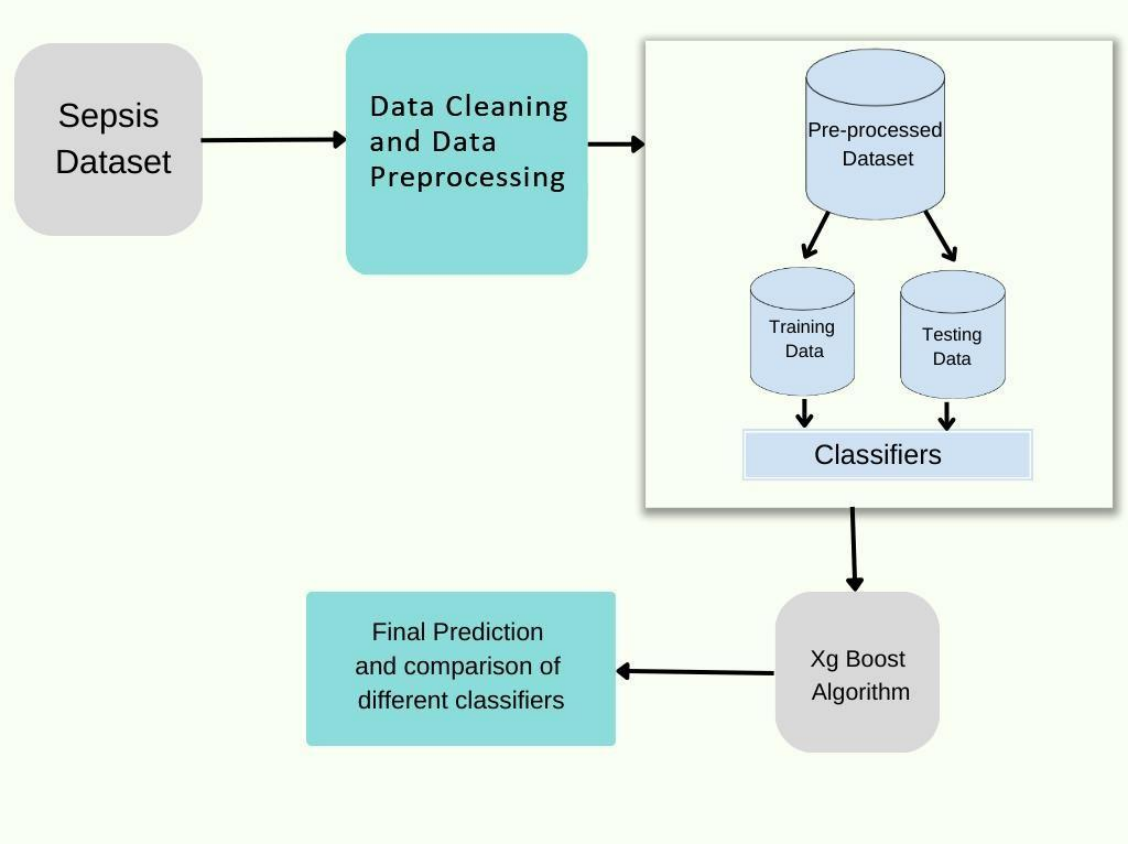


Fig 1.1 - Architecture

Working of Architecture -

1. Obtain the sepsis dataset.
2. Clean and preprocess the dataset to handle missing values and outliers.
3. Split the preprocessed data into training and testing sets for model evaluation.
4. Utilize various classifiers such as Random Forest, Naive Bayes, and KNN to train on the training data.
5. Enhance model performance by using the XG Boost algorithm.
6. Finally, make predictions based on the combined performance of the different classifiers, potentially employing ensemble methods for improved accuracy and robustness in sepsis detection.

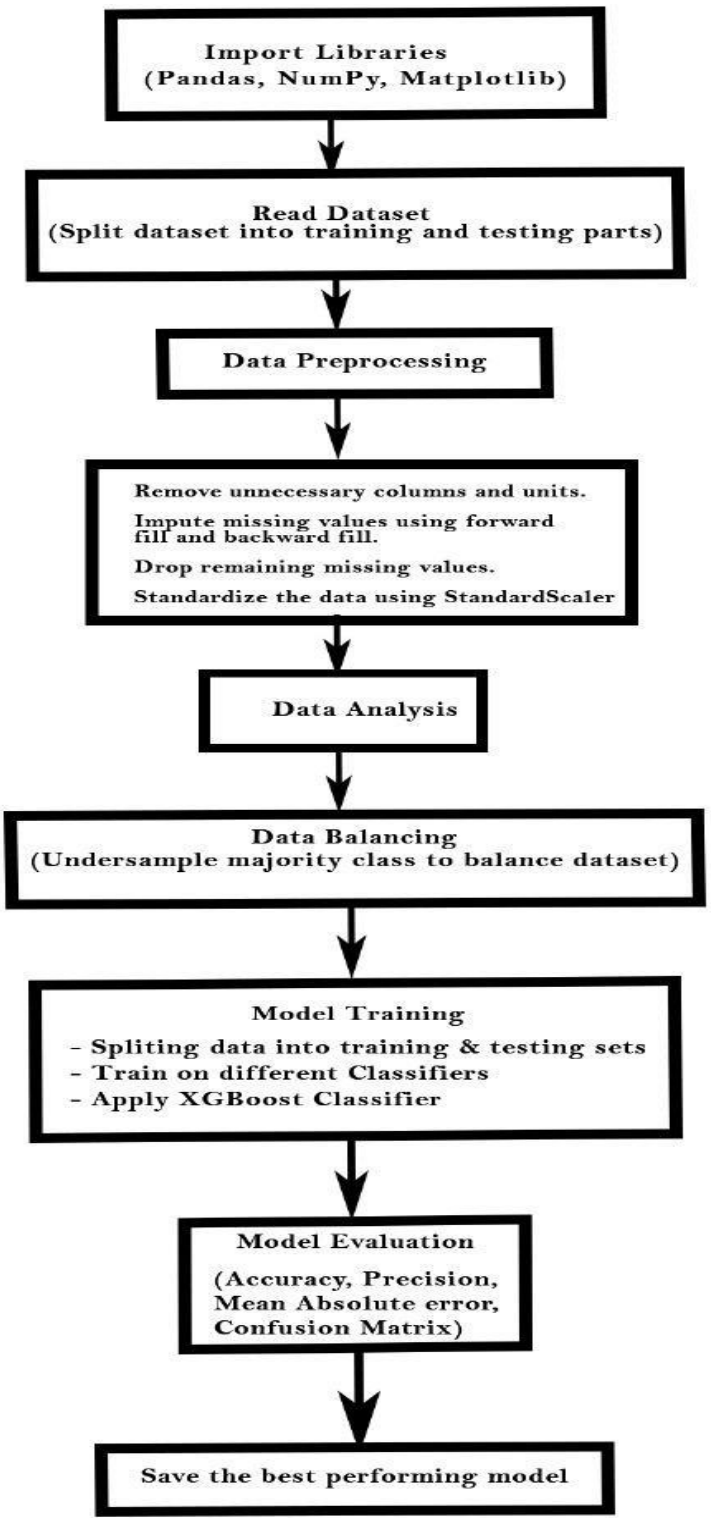


Fig 1.2- Working Flowchart

Working of Flowchart -

1. Data is prepared by importing libraries, reading data from Google Drive, and pre-processing it by removing unnecessary columns and handling missing values.

2. Analysis is conducted to understand feature correlations. Balancing techniques are applied to address any class imbalances in the data set.

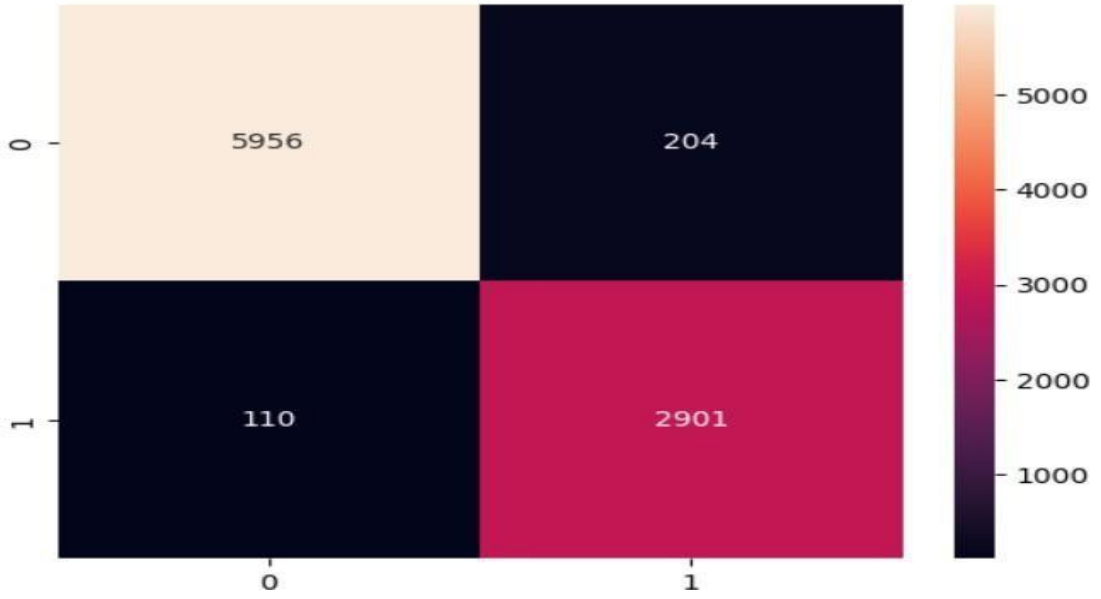
3. Multiple classification models like Random Forest and Logistic Regression are trained on the prepared data.

4. Evaluation metrics such as accuracy and F1 score are used to assess model performance. Finally, the best performing model is selected and saved for future use.

5. RESULTS

To summarize, in this project, we have created a web interface that can be used by doctors and healthcare communities which will help them to detect sepsis at an early stage. We also have looked and compared different algorithms that are used for detecting sepsis. In this study, among all the classifiers that were considered for detection of sepsis infection on comparing them randomforest was found as the best classifier as compared to others based on performance measures. It has an accuracy of almost 96 percent. The performance measures of random forest are shown in below figure :-

```
Accuracy: 0.9657616399520227
Precision: 0.9342995169082126
Recall: 0.9634672866157423
F1 Score: 0.9486592544146502
AUC-ROC: 0.9651752017494296
Mean Absolute Error: 0.03423836004797732
Root Mean Squared Error: 0.1850361047146673
```



Based on the comparison between different classifiers done in this study we will say that the random forest is found as best to detect sepsis efficiently and with accuracy.

6. CONCLUSION AND FUTURE WORK

The paper has proposed the idea to make a web interface that will serve people working in various health care communities for predicting sepsis. Every health care organization will have the privilege of treating sepsis with no long lasting effects and with almost a full recovery that will ultimately benefit both patients and doctors. This system has presented a description about Sepsis. The symptoms of the disease, signs, complications, and early diagnose for the disease are presented. Sometimes it is very challenging for doctors to detect sepsis at an early stage because sepsis is not just a regular infection but rather complex life threatening condition also the symptoms of sepsis may overlap with some existing medical conditions Symptoms of this disease may progress swiftly from mild to severe stage which can make detecting sepsis at an early stage challenging for healthcare providers.

So in this study, we have explored various machine learning algorithms for sepsis detection using data from the Physio Net Challenge dataset. Among all the algorithms we used for detection of sepsis, Random Forest was found as the most accurate method for predicting sepsis with an accuracy of approximately 97 percent better than all the other machine learning algorithms that we have compared in this study. The identification of Random Forest as the most effective algorithm for sepsis detection holds significant implications for clinical practice and will prove to be an invaluable tool in future for detecting sepsis at an early stage by making more advancements in this model with time. This system will help healthcare providers for early diagnose of the disease which will ultimately help to lower the mortality and morbidity rate that are caused due to sepsis. This approach will not only help doctors to efficiently predict sepsis but also will act as a common platform for all health care organizations to take it in their practical usage for predicting sepsis at an early stage and improve the patient outcomes.

This interface will have a very high future scope because it can serve the emergent needs of doctors to predict sepsis accurately at an early stage. Collaborating with healthcare providers for real world testing and making changes accordingly before getting it deployed as website in hospitals to improve the overall accuracy and reliability of the system. In future, we would like to enhance the application by making it more customer specific and deploying this model in hospital websites and help the doctors to detect early signs of the disease and also by predicting which type of sepsis and how much the patient is affected by the disease. Collaborate with medical professionals for more deep research, to ensure a better understanding of sepsis and other related information which is critical for them related to sepsis to make advancements in the accuracy and usability of build predictive models can help to get these systems deployed for real world application.

7. REFERENCES

[01] - Baldominos, A., Puello, A., O˘gul, H., A¸suro˘glu, T., Colomo- Palacios, R. (2020). Predicting infections using computational intelligence–a systematic review. IEEE Access, 8, 31083-31102.

[02] - Al-Mualemi, B. Y., Lu, L. (2020). A deep learning-based sepsis estimation scheme. IEEE Access, 9, 5442-5452.

[03] - Feng, M., McSparron, J. I., Kien, D. T., Stone, D. J., Roberts, D. H., Schwartzstein, R. M., ... Celi, L. A. (2018). Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database. Intensive care medicine, 44(6), 884-892.

[04] - Kam, H. J., Kim, H. Y. (2017). Learning representations for the early detection of sepsis with deep neural networks. Computers in biology and medicine, 89, 248-255.

[05] - Wang, X., Wang, Z., Weng, J., Wen, C., Chen, H., Wang, X. (2018). A new effective machine learning framework for sepsis diagnosis. IEEE access, 6, 48300-48310.

[06] - Zhao, X., Shen, W., Wang, G. (2021). Early prediction of sepsis based on machine learning algorithm.Computational Intelligence and Neuroscience, 2021.

[07] - Wang, Z., Yao, B. (2021). Multi-branching temporal convolutional network for sepsis prediction. IEEE Journal of Biomedical and Health Informatics, 26(2), 876-887.

[08] - Islam, M. M., Nasrin, T., Walther, B. A., Wu, C. C., Yang, H. C., Li, Y. C. (2019). Prediction of sepsis patients using machine learning approach: a meta-analysis. Computer methods and programs in biomedicine, 170, 1-9.

[09] - Fu, M., Yuan, J., Lu, M., Hong, P., Zeng, M. (2019, September). An ensemble machine learning model for the early detection of sepsis from clinical data. In 2019 Computing in Cardiology IEEE.

[10] - Ying, T. X., Abu-Samah, A. (2022, June). Early Prediction of Sepsis for ICU Patients using Gradient Boosted Tree. In 2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS) (pp.78-83).

[11] - Liu, S., Fu, B., Wang, W., Liu, M., Sun, X. (2022).Dynamic sepsis prediction for intensive care unit patients using XG Boost-based model with novel time-dependent features.IEEE Journal of Biomedical and Health Informatics.

[12] - Scherpf, M., Gr¨aßer, F., Malberg, H., Zaunseder, S. (2019).Predicting sepsis with a recurrent neural network using the MIMIC III database. Computers in biology and medicine, 113, 103395.