ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

VITAL SEPSIS: EARLY PREDICTION OF SEPSIS

Bhandaram Snehasriya[1], Pandagre Ajay[2], Vanjari Nissangi Vaibhav[3], Dr.I.Nagaraju[4]

[1,2,3]B.Tech Student, Department of CSE (Data Science), Malla Reddy College of Engineering and Technology, Hyderabad, India.

[4] Professor, Department of CSE (Data Science), Malla Reddy College of Engineering and Technology, Hyderabad, India.

***Abstract*—** Sepsis is a life-threatening condition requiring early detection for effective treatment. Machine learning techniques have emerged as valuable tools for predicting sepsis onset and improving patient outcomes. This abstract summarizes the current state of research on sepsis prediction using machine learning. Machine learning models utilize a variety of clinical data, including vital signs, laboratory results, medical history, and demographics, to generate accurate predictions. Various algorithms, such as logistic regression, decision trees, random forests, support vector machines, and deep learning, have been employed to develop sepsis prediction models. Data preprocessing techniques, including feature selection and normalization, enhance the quality of input data. Incorporating dynamic and temporal features captures the evolving nature of sepsis progression. Evaluation of sepsis prediction models involves retrospective clinical datasets and performance metrics such as sensitivity, specificity, AUC-ROC, and F1 score.

Keywords— sepsis; early detection and treatment; generalizability; open-source algorithms; competition; sequential prediction tasks

I. INTRODUCTION

Sepsis is a severe and potentially life-threatening medical condition that arises when the body's response to an infection becomes dysregulated, leading to widespread inflammation, organ dysfunction, and, in severe cases, death. Timely intervention is crucial to improving outcomes for septic patients, making early detection a cornerstone of effective sepsis management. Vital sepsis disease prediction, a burgeoning field at the intersection of medicine and technology, aims to harness the power of data and predictive analytics to identify patients at risk of developing sepsis before clinical symptoms become apparent.

In recent years, there has been a growing recognition of the need for more proactive approaches to sepsis management. Healthcare systems have increasingly adopted electronic

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

health records (EHRs) and continuous monitoring devices, generating vast amounts of patient data. Vital signs, laboratory results, and patient history are among the key data points that can provide valuable insights into a patient's health status. Leveraging this wealth of information, researchers and clinicians are now employing advanced machine learning algorithms and artificial intelligence techniques to develop predictive models that can forecast sepsis onset.

This introduction sets the stage for an exploration of vital sepsis disease prediction, highlighting the pressing need for early sepsis detection and the potential of data-driven predictive models to revolutionize sepsis management. By combining the expertise of medical professionals with the capabilities of modern technology, vital sepsis prediction holds promise in enhancing patient care, reducing healthcare costs, and ultimately saving lives in the ongoing battle against sepsis.

The reliable and early identification of sepsis is often complicated by its syndromic nature, which can contribute to delays in treatment. The importance of early identification and treatment of sepsis is highlighted in two recent studies that suggest an increase in the adjusted mortality of septic patients who experienced delays in antibiotic therapy. This effect is even more profound in patients suffering from septic shock, where hourly delays were associated with an 3.6-9.9% increase in mortality per hour. Professional critical care societies have proposed clinical criteria for recognizing and treating sepsis; however, the fundamental need for early and reliable identification of sepsis remains unmet.

In this paper, we begin with the Challenge objective of early predictions of sepsis, the Challenge data and clinical criteria for sepsis, and a new evaluation metric that reflects 4 the clinical utility of early sepsis predictions. We continue with the Challenge submission procedure, the results of the Challenge, and a discussion of computational approaches for early predictions of sepsis.

II. LITERATURE REVIEW

The use of prior data to estimate the pricing of available and new launch products is an intriguing study background for machine-learning researchers. Sameerchand-Pudaruth[1] estimates the prices of used automobiles in Mauritius. He used a variety of approaches to forecast prices, including multiple linear regressions, k-nearest neighbours (KNN), Decision Tree, and Nave Bayes. Sameerchand-Pudaruth obtained equivalent results using all of these approaches. During study, it was discovered that the majority of standard algorithms, such as Decision Tree and Nave Bayes, are incapable of processing, categorising, and forecasting numeric data. There were only 97 instancesof his work (47 Toyota+38 Nissan+12 Honda). Because of the small number

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

of cases used, the forecast accuracies were quite low [1]

Shonda Kuiper [2] was also employed in the same profession. Kuiper utilised a multivariate regression model to predict the pricing of 2005 General Motors vehicles. He gathered the information from www.pakwheels.com, an internet resource. The "Introduction of acceptable approaches for variable selection," which assisted in determining which variables are more suitable and relevant for model inclusion, is the main component of this study effort. This (His research) enables students and potential researchers in a variety of areas to examine the conditions under which investigations should be conducted and to identify when appropriate approaches should be used[2].

Mariana Listiani[3], another researcher, uses the notion of supporting vector machine (SVM) for the same study. Listiani utilised the aforementioned method to anticipate leased car costs. When there is a big data set, it was discovered that the SVM approach is considerably better and more dependable for price prediction than other methods such as multiple linear regressions. SVM is also excellent at managing high-dimensional data and minimising both under-fitting and over-fitting issues, according to the study. Essential characteristics for SVM Listiani were identified via a genetic algorithm. In terms of variance and mean standard deviation, however, the approach failed to show why SVM is superior than basic multiple regression [3].

The Limsombunchai research [4] revealed that neural networks (NN) are effective at estimating property prices. When compared to the hedonic technique, his method was more dependable. Except for the fact that the model is first trained in NN and then assessed for prediction, all techniques function the same. NN produced greater R-sq and lower root mean square error (RMSE) using both approaches, but the hedonic offered lower values. The study was restricted since real housing values were unavailable, and the analysis work relied solely on estimates[4].

K Noor and SadaqatJ[5] have experimented with a variety of methods for estimating car prices. The researchers attained optimum accuracy by using several linear regressions. This article offers a method in which pricing is anticipated based on variables such as car type, make, area, edition, colour, mileage, alloy rims, and power steering [5].

III. METHODOLOGY

A. Data Collection:

We started by collecting data from various sources. three geographically distinct U.S. hospital systems with three different electronic medical record systems: Beth Israel Deaconess

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

Medical Center (hospital system A), Emory University Hospital (hospital system B), and a third, unidentified hospital system (hospital system C). These data were collected over the past decade with approval from the appropriate Institutional Review Boards. We de-identified and labeled the data using Sepsis-3 clinical criteria .

Data and labels for 40,336 patients from hospital systems A and B were posted publicly for download and data and labels for 24,819 patients from hospital systems A, B, and C were sequestered as hidden test sets. Datasets like MIMIC-III and eICU Collaborative Research Database are valuable resources for sepsis prediction.

B. Data Preprocessing:

Clean the dataset by handling missing values, outliers, and irrelevant columns. Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding. Split the dataset into training and testing sets.

C. Feature Selection:

Determine which features are most relevant for sepsis prediction. You can use statistical methods or machine learning algorithms like Random Forest for feature selection.

	Measurement	Description
1	HR	Heart rate (beats per minute)
2	O2Sat	Pulse oximetry (%)
3	Temp	Temperature (deg C)
4	SBP	Systolic BP (mm Hg)
5	MAP	Mean arterial pressure (mm Hg)
6	DBP	Diastolic BP (mm Hg)
7	Resp	Respiration rate (breaths per minute)
8	EtCO2	End tidal carbon dioxide (mm Hg)
9	BaseExcess	Excess bicarbonate (mmol/L)
10	HCO3	Bicarbonate (mmol/L)
11	FiO2	Fraction of inspired oxygen (%)
12	pH	pH
13	PaCO2	Partial pressure of carbon dioxide from arterial blood (mm Hg)
14	SaO2	Oxygen saturation from arterial blood (%)
15	AST	Aspartate transaminase (IU/L)
16	BUN	Blood urea nitrogen (mg/dL)
17	Alkalinephos	Alkaline phosphatase (IU/L)
18	Calcium	Calcium (mg/dL)
19	Chloride	Chloride (mmol/L)
20	Creatinine Creatinine	(mg/dL)
21	Bilirubin_direct	Direct bilirubin (mg/dL)
22	Glucose	Serum glucose (mg/dL)
23	Lactate	Lactic acid (mg/dL)
24	Magnesium	Magnesium (mmol/dL)
25	Phosphate	Phosphate (mg/dL)
26	Potassium	Potassiam (mmol/L)
27	Bilirubin_total	Total bilirubin (mg/dL)
28	TroponinI	Troponin I (ng/mL)
29	Hct	Hematocrit (%)
30	Hgb	Hemoglobin (g/dL)
31	PTT	Partial thromboplastin time (seconds)
32	WBC	Leukocyte count (count/L)
33	Fibrinogen	Fibrinogen concentration (mg/dL)
34	Platelets	Platelet count (count/mL)
35	Age	Age (years)
36	Gender	Female (0) or male (1)
37	Unit1	Administrative identifier for ICU unit (MICU); false (0) or true (1)
38	Unit2	Administrative identifier for ICU unit (SICU); false (0) or true (1)
39	HospAdmTime	Time between hospital and ICU admission (hours since ICU admission)
40	ICULOS	ICU length of stay (hours since ICU admission)
41	SepsisLabel	For septic patients, SepsisLabel is 1 if $t \geq t_{\text{sepsis}} - 6$ and 0 if $t < t_{\text{sepsis}} - 6$. For non-septic patients, SepsisLabel is 0.

D. Model Selection:

Choose an appropriate machine learning model. Common models for sepsis prediction include logistic regression, decision trees, random forests, support vector machines, and neural networks.

E. Evaluation Metrics:

Use appropriate evaluation metrics such as accuracy, sensitivity, specificity, area under the ROC curve (AUC), F1-score, and precision-recall curves to assess model performance.

Implement cross-validation to ensure robust model evaluation.

F. Data Mining:

We tested different prediction models in the data mining phase, such as logistic, SMO,

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

naïve Bayes, JRip, and KNN. We also used proportional hazard regression for survival analysis. We classified individuals as having sepsis (1) or not (0). Survival analysis involves analyzing data in which the outcome variable is the time until a specific event occurs. Proportional hazard regression is the most widely used method for studying how predictor variables impact survival time. Logistic regression classifies observations using a model and probability estimates, which is effective for categorical data. In machine learning, support vector machines (SVM) are models that can tackle classification or regression challenges in a supervised learning setting. Among the types of SVM is sequential minimal optimization (SMO), which is particularly efficient in solving classification problems and handling regression issues. KNN is a simple algorithm for regression and classification in machine learning. It divides the dataset into classes for prediction. The attributes were selected by examining the literature and getting input from intensive care doctors. Although a high C-reactive protein (CRP) level can indicate an infection, it may not be precise enough to identify sepsis. It is advisable to include more biomarkers in upcoming research to enhance the early prognosis for sepsis patients. The study includes ten attributes: gender, time, blood sugar, lactic acid, SBP, SpO2, heart rate, white blood cell count, temperature, and respiratory rate. We conducted the experiments using split data and internal cross-validation with ten folds.

IV. SYSTEM DESIGNS

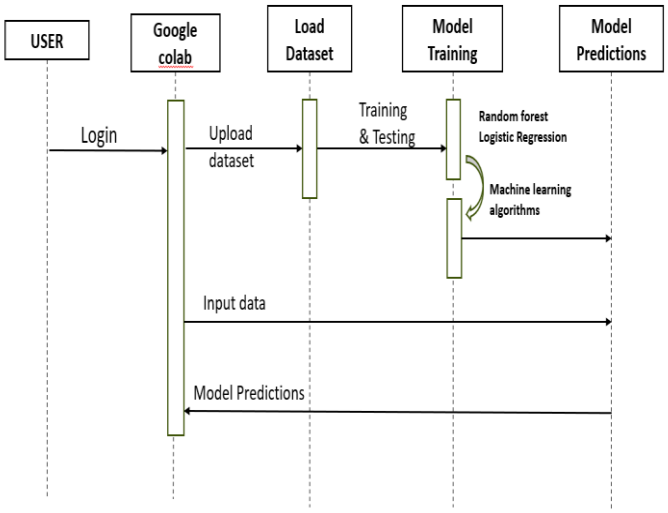

Fig: Dataflow Diagram

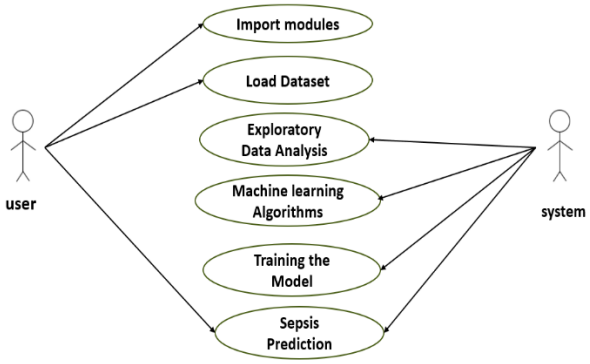

Fig: Use case Diagram

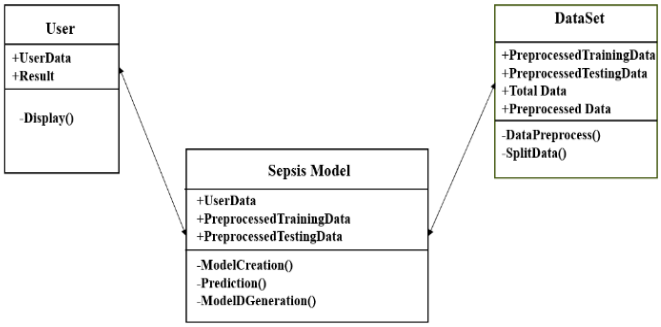

Fig: Class Diagram

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

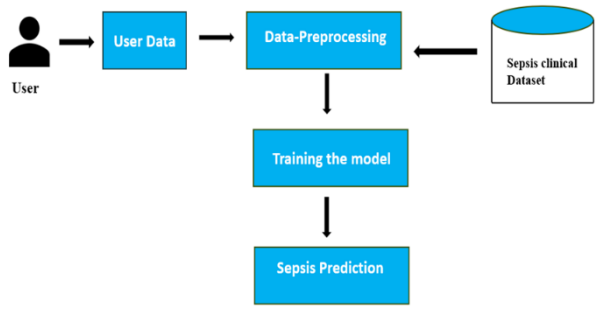

Fig: Architecture Diagram

V. RESULTS

We conducted a thorough analysis of the medical records of a total of 1182 patients who were diagnosed with sepsis. In this study, we explored two approaches for predicting sepsis in its early stages. The first technique uses a regression model known as survival analysis, and the second uses a data mining algorithm. Our dataset comprises valuable information on vital signs, lab tests, and demographic characteristics. However, the dataset is primarily dominated by negative instances, with only a few positive ones. This class imbalance poses a significant challenge for data mining because most algorithms focus on classifying large samples and disregarding minority samples. Nonetheless, in some cases, the minority samples are critical for accurate predictions. To overcome this issue, we opted to use the under-sampling technique. Metric should be considered for wider adoption in clinical care because it does not suffer from many of the problems of F-measures (and related metrics such as accuracy, sensitivity, and positive predictive value) or standard area under the curve metrics (such as AUROC and AUPRC), which either assume a one-shot decision or no decision threshold.

VI.CONCLUSION

Early detection of sepsis is crucial in healthcare, and recent research has explored new ways to predict its onset accurately. Researchers have analyzed data from one tertiary hospital database using advanced machine learning algorithms and survival analysis. Based on our research, machine learning models are better at predicting sepsis in patients than traditional methods. However, the model is more reliable at identifying patients who do not have sepsis than those who do, which is consistent with other studies. Our study emphasizes that sepsis can be anticipated by three crucial factors: time, temperature, and lactic acid levels. Although the results were moderate, efforts are underway to improve the models by applying more data preprocessing and attribute selection. The study will expand to settings beyond intensive care units and extend the analysis period. These efforts will help enhance future models and improve early detection of sepsis, ultimately saving more lives.

VII. REFERENCES

[1] Christopher W. Seymour, Vincent X. Liu, Theodore J. Iwashyna, Frank M. Brunkhorst, Thomas D. Rea, Andr´e Scherag, Gordon Rubenfeld, Jeremy M. Kahn, Manu ShankarHari, Mervyn Singer, et al. Assessment of clinical

ISSN2454-9940
www.ijsem.org
Vol 18, Issuse.1 March 2024

criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Journal of the American Medical Association, 315(8):762–774, 2016.

[2] Mervyn Singer, Clifford S. Deutschman, Christopher Warren Seymour, Manu ShankarHari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R. Bernard, Jean-Daniel Chiche, Craig M. Coopersmith, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Journal of the American Medical Association,315(8):801–810,2016.

[3] Manu Shankar-Hari, Gary S Phillips, Mitchell L Levy, Christopher W Seymour, Vincent X Liu, Clifford S. Deutschman, Derek C. Angus, Gordon D. Rubenfeld, and Mervyn Singer. Developing a new definition and assessing new clinical criteria for septic shock: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Journal of the American Medical Association, 315(8):775–787, 2016.

[4] Centers for Disease Control and Prevention. Sepsis. https://www.cdc.gov/sepsis/datareports/index.html, 23 August 2016. [Online; accessed 1 February 2019].

[5] World Health Organization. Sepsis. https://www.who.int/news-room/fact-sheets/detail/sepsis, 19 April 2018. [Online; accessed 1 February 2019].

[6] Manu Shankar-Hari, Gary S Phillips, Mitchell L Levy, Christopher W Seymour, Vincent X Liu, Clifford S. Deutschman, Derek C. Angus, Gordon D. Rubenfeld, and Mervyn Singer. Developing a new definition and assessing new clinical criteria for septic shock: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Journal of the American Medical Association, 315(8):775–787, 2016.

[7] Centers for Disease Control and Prevention. Sepsis. https://www.cdc.gov/sepsis/datareports/index.html, 23 August 2016. [Online; accessed 1 February 2019].

[8] World Health Organization. Sepsis. https://www.who.int/news-room/fact-sheets/detail/sepsis, 19 April 2018. [Online; accessed 1 February 2019].

[9] Greg S. Martin, David M. Mannino, Stephanie Eaton, and Marc Moss. The epidemiology of sepsis in the United States from 1979 through 2000. New England Journal of Medicine, 348(16):1546–1554, 2003.

[10] J.L. Moran, J.A. Myburgh, G.A. Syres, D.A. Jones, P.A. Cameron, A. Higgins, S. Finfer, S. Webb, A. Delaney, A. Cross, et al. The outcome of patients with sepsis and septic shock presenting to emergency departments in Australia and New Zealand. Critical Care and Resuscitation, 9(1):8, 2007.

[11] Jeremy Stoller, Laura Halpin, Matthew Weis, Brett Aplin, Weikai Qu, Claudiu Georgescu, and Munier Nazzal. Epidemiology of severe sepsis: 2008-2012. Journal of Critical Care, 31(1):58–62, 2016.