Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

# A Comprehensive Review on Predictive Analytical Solutions for Sepsis Patients

**[1*]Smitha N, [2]Tanuja R, [3]S H Manjula, [4]Venugopal K R**

[1]Research Scholar, Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore , 560001, Karnataka, India.
[2]Department of Computer Science and Engineering, CMR Institute of Technology, Bangalore, Karnataka, India.

**Abstract**— Sepsis is a grave medical complication that arises from an infection within the body, with the potential to result in tissue damage, organ dysfunction, or even fatal outcomes. Early detection is very much required to reduce the mortality rate. The different types of sepsis-like Sepsis-1, Sepsis-2, and Sepsis-3. The research community has a wide range of study options since the definition of sepsis is always changing. The gold standard for early sepsis diagnosis has been used in research about sepsis detection, although it performs poorly in comparison to machine learning algorithms. In this paper a detailed review of research papers is done to do an analysis of data source, feature selection, metric selection, and future scope

**Keywords—** CNN, Deep Neural Network, Infection, Mortality rate, Machine learning, Multiple organ failure, Neonate, RNN, Sepsis, Septic Shock, Severe Sepsis.

## 1. INTRODUCTION

Sepsis, a potentially fatal medical illness, develops as a result of an infection in the body. It represents a critical and complex challenge in healthcare, as it can rapidly progress and lead to severe complications, including organ dysfunction and death.

Sepsis can affect individuals of all ages, from newborns to the elderly, and it remains a significant global health concern.

Sepsis happens when the body's immune system, which typically responds to infections by fighting off the invading pathogens, goes awry as shown in Fig1. Instead of containing the infection, the immune response triggers widespread inflammation throughout the body. This excessive inflammation can result in a cascade of events, including blood clotting, blood vessel dilation, and reduced blood flow to vital organs.

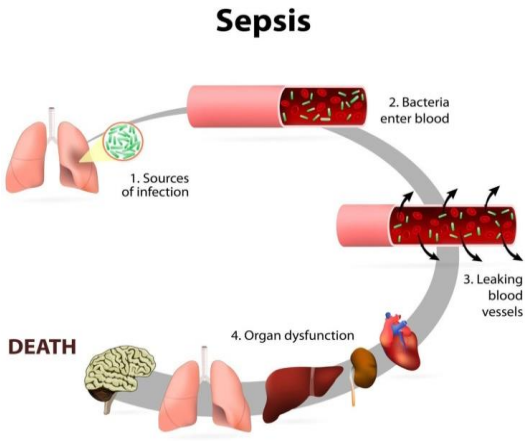


**Fig. 1** Septic Stages Timeline[1]
https://www.news-medical.net/health/The-Stages-of-Sepsis.aspx

Timely diagnosis and intervention are crucial in managing sepsis effectively. However, sepsis is often challenging to identify early, as its symptoms can be nonspecific and mimic other common illnesses. This complexity in diagnosis and the urgency of treatment make sepsis a pressing issue in healthcare, prompting ongoing research and efforts to improve early detection and treatment strategies.

There are different definitions of Sepsis from 1992 to 2001. It is defined as a syndrome response to infection resulting in inflammation as shown in Table 1.

People who are hospitalized or have recently been hospitalized or People in ICU are more vulnerable to developing infection, which could result in sepsis [1].

Sepsis is basically diagnosed based on blood tests to check for infection, clotty problems, organ functionality, decreased oxygen amount, and imbalance in acidity in the blood. A urine test is conducted to check bacteria that may lead to infection, a wound secretion test, mucus secretion test. If it cannot be done for doctors to find the source of infection using these tests then interior perspective of the body is done using a CT-Scan, x-ray, ultrasound, and MRI scan.

The source of infection using these tests then interior perspective of the body is done using a CT-Scan, X-Ray, Ultrasound, and MRI Scan.

**Table 1:** Stage in Sepsis

| **Sepsis** | **Severe Sepsis** | **Septic Shock** |
|---|---|---|
| Fever>108 F(38 c) <96.8 F(36 C) | Organ failure, Discolored skin | Includes symptoms of Severe Sepsis plus |
| Heart rate >90 beats/min | Decreased urination, change in mental ability | A very low blood pressure |
| Breathing rate higher than 20 br/min | Low platelet count | |
| Probable or confirmed infection | Chills, weakness, unconsciousness | |

## 2. Literature Review

**Corrnee [2]** et al., used the NSQIP database to extract patients suffering appendectomy between 2015-2017 and Machine learning algorithms used to predict postoperative sepsis.

**T.E.Garcia Gallo [3]** et.al., created a model to predict 1-year death in sepsis- diagnosed critical patients. Here data are extracted from the MIMIC III database of 5650 admissions of patients with sepsis. The basis for this concept is the Stochastic Gradient Boosting methodology. Performance was evaluated using AUROC and 0.8039 was obtained. Here patients with severe sepsis or septic shock are selected. The limitation is model cannot be used after 24 hours of observation and data is used only from one institution.

**Davide[4]** et al., have examined a dataset of 110,204 hospital admissions of 84,811 patients in Norway between 2011 and 2012. These patients were diagnosed with infections, SIRS, Sepsis, or Septic shock. The authors have selected three characteristics of patients with sepsis that were noted at the time of admission: sex, age, and number of septic episodes. The AUC obtained was 0.86 and the model was tested with 137 patients from an external resource and obtained AUC=0.863. The study's limitations are, given that this methodology is successful in identifying patients who have been admitted yet are still living, and the same cannot be stated for the admissions of patients who have passed away. Time-related forecasts, which would have a greater impact in a therapeutic situation, are suggested as future studies.

**Andres[5]** et al., have evaluated a supervised learning algorithm for the prediction of mortality. Data are analyzed from 3 university Hospitals in Medellin, Colombia. Here patients who have encountered infection, suspected and any organ dysfunction are included. Here Decision tree, Random Forest, ANN SVM model is used. The mortality rate is 11.5% and SVM results in AUROC of 0.69 and ANN of 0.69. The advantage of this study is that heterogeneous institutes are considered for data collection and cross-validation carried out and obtained consistent results.

**Armando[5]** et al., locally cured EHR data. Here authors have compared three ML models Random forest, Cox regression penalized, logistic regressions. Here MGR- RNN outperforms within 4h of onset with C-statistic off. Baseline features considered are Age, sex, weight, admission type, race, comorbidities, etc. 42,979 admissions were considered where 8160 admissions developed sepsis. Limitations of this model are authors have not included acute respiratory dysfunction. A geographically separate population is not used to test the created model.

## 3. Background Work

Sepsis is a serious issue for global health. By carrying out prompt interventions, timely treatment improves mortality. [6].Prior research on hospital administrative datasets was used to estimate sepsis incidence

and death. That only contained information on patients who were admitted [7]. Other research has utilized electronic health record data [8] and death certificates [9].

There are three stages in Sepsis, they are Sepsis, Severe Sepsis, and Septic Shock[10] as shown in Fig 2. Where Sepsis manifests few symptoms like temperature>38 C or 108 F and <96.8 F or 36 C, heart rate greater than 90 beats/min, breathing rate higher than 20 breaths per min, WBC count >12,000 per ml or lesser than 4000 per ml and probable or confirmed infection [11]. Severe sepsis is associated with a few symptoms like organ failure, decreased urination, changes in mental ability, chills, unconsciousness, weakness, etc [12]. All the signs of severe sepsis are present in septic shock, in addition to very low blood pressure as shown in Table 1.

Mortality From 2009 to 2019, investigators looked at the death rates of individuals with sepsis and septic shock in many regions. That has done a comparison between 30 and 90 days mortality.

Who is at risk?

- Anyone affected by an infection
- Severe injury
- Non-communicable disease
- Older persons
- Pregnant women
- Neonates
- Hospitalized patients

**Table 2:** Comparison of Advantages, Disadvantages, and Challenges of Work in Predictive Analytic of Sepsis

| **Category** | **Approaches** | **Advantages** | **Disadvantages** | **Challenges** |
|---|---|---|---|---|
| *Sepsis-1 Simo et al* | Deep learning approach in a heterogenous data set outside ICU for early detection of sepsis | Doesn't restrict on the set of clinical parameter and on the cure of sepsis | Conclusion are based on hypothesis-generating nature | Clinical acceptance by including prospective conforming study |
| *Sepsis-3 Christopher et al* | Deep learning Temporal Neural Network | Sizeable data used for training, model validated by 3 technique hence robust, data from two distinct clinical contexts are included in the unlocalized data set, and feature extraction and selection were automatically performed by the model. | GPR system is time-consuming, the output produced is high variance and noise, the compilation is expensive | Clinical acceptance by including prospective conforming study |
| *Sepsis-1 Christopher Baston et al* | Deep learning approach | Can Detect sepsis up to 48 h in advance of onset | AUROC, Sensitivity are calculated using the smaller patient pool with longer prediction window | Training using small training data set |
| *Sepsis-1 Mathieu et al* | Predicting Sepsis 3 hr prior to sepsis onset and also for 6 and 12 h | using RNN fine-tuning of the model to different data is done | Black box character, specificity is 47% of the classifier | Using database which is biased for specific country and region |

4. DATA COLLECTION SOURCES

Data are gathered from many sources. Many researchers have considered multiple sources databases to carry out the predictive analysis. Christopher et al., have used data from the University of California San Francisco (UCSF) Medical Center and the Beth Israel Deaconess Medical Center (BIDMC). Most researchers have employed MIMIC III database data, which is biased towards a certain location.

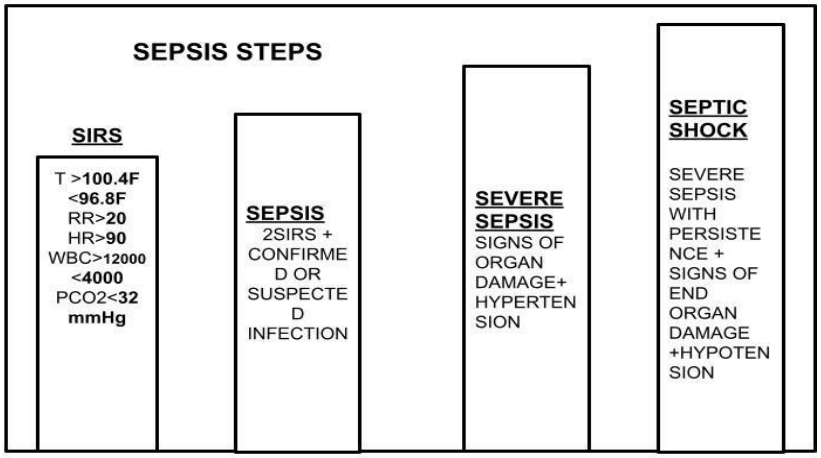


**Fig. 2** Sepsis Steps[2]

Christopher used an open-source data set released for the PhysioNet Computing in Cardiology 2019 challenge which contained 2 hospitals Beth Israel Deaconess Medical Center and Emory University Hospital with 1790 and 1142 septic patient records [13]. Data gathered from two databases were utilized in two investigations, one from the University of California, San Francisco (UCSF) Medical Center and Beth Israel Deaconess Medical Center (BIDMC). Davide [4] et al., have analyzed a dataset of 110,204 admissions of 84,811 hospitalized subjects between 2011-2012 in Norway, these patients were diagnosed with infections, SIRS, Sepsis, or septic shock. Andres et al used 3 university hospitals in Medellin, Colombia, where people who experienced infection are included in this group of patients, suspected or any organ dysfunction. Few studies have taken locally cured EHR data of different age groups, starting from neonates to adults to create a model and assess it against other ML models to find the best to predict sepsis as shown in Table 2.

### A. Variability in Feature selection:

Features selected for predicting sepsis have a wide range. Most of the features selected are vital signs, laboratory measurements, demographic variables, Bacterial infection, fungal infection, organ dysfunction, sex, age, specific episodes, partial pressure of oxygen, decreased Glasgow coma score(GCS), weight, admission type, race, comorbidities, gene expression profiles, platelet counts, gender, birth weight, band cells, catheter used, hematology, clinical data like date of admission, diagnosis, HRV features, blood pressure, respiratory rate, a primary score of infection as shown in Table 2.

### B. Feature Engineering:

Christopher et al., used two methods to fill in missing values. 1)Gaussian Process Regression(GPR):- Used to lessen the problem of missing values by predicting the distribution of possible values for each attribute contained entries. 2) Radial Basis Function (RBF):- White noise and a combined kernel are utilized to create a covariance matrix that depicts the distribution values.

### C. Feature Selection:

The Patients with 8 vital signs, 26 lab measurements, and 6 demographic factors that were recorded hourly are selected as features [14]. Prediction is done 24 hr and 48 hr prior to sepsis onset and achieved good performance. Somanch et al., extracted 30 h clinical data from the MIMIC III database to train a model where the model predictscardiac arrest 6h prior with good accuracy as shown in Table 3.

### D. Random Forest:

In previous works, different types of classification algorithms used are briefly discussed in this section. Random Forest consists of individual decision trees that operate on an ensemble technique.

Another name for Random Forest is Bootstrap aggregation. They construct unique trees to randomly sample from the dataset with replacement, producing various trees as a consequence. This procedure is referred to as bagging. Now, utilize the Gini value or Entropy to determine the best-split impurity and information gain for each tree to find the best one [15].

### E. Extra Gradient boost:

XGboost can be applied to classification or regression predictive modeling issues. This algorithm relies heavily on a decision tree-based boosting strategy. Its speed and strong model performance are advantages. We can adjust hyper parameters like tree depth, learning rate, sample size, and feature count [16].

### F. SVM:

The SVM algorithm's goal is to categorize the data points by locating a suitable hyper plane in a space. The major goal is to find the maximum hyper plane. The bounds of hyper planes serve to categorize the data points. The margin between the data points and the hyper plane must be maximized in the SVM method [17].

**Table 3:** Comparison of paper with respect to population location, feature set size, age group type of data

| Authors | Year | Population Location Size | Feature set size | Age | Type of data |
|---|---|---|---|---|---|
| *Simon et* | 2020 | Multicenter data set from outside ICU, Danish Hospital dataset, EHR data combined with data from National Patient Register and One civil Registration System. | Danish hospital collected data over years(2010-2017), Data extracted from research projects 'CROSS-TRACKS' | >18 Years | EHR includes Biochemistry, medicine, microbiology, medical image and the patient administration system(PAS), NPR and CRS includes Contextual type like registered to diagnose, procedures, hospital admissions, marital status, housing situation |
| *Christopher et al* | 2020 | Open Source dataset from PhysioNet Computing in cardiology Beth Israel, Deaconess, Medical center, Emory university hospital | 1790 Septic Records | >18 Years | 40 features,8 vital signs,26 laboratory measures,6 demographic variables, recorded hourly |
| *Christopher Barton et al* | 2019 | Database from the University of California, San Francisco (USCF) Medical Center and the Israel Deaconess Medical Center (BIDMC) | UCSF dataset 17,467,987 inpatient and outpatient and BIDMC dataset | >=18 Years | 6 Vital Signs like systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, peripheral Oxygen saturation, and temperature |

## 5. Performance Metrics

Different performance metrics were used in previous work to analyze the performanceof the model built. Briefly, they are listed below.

### A.AUC-ROC:

The area under Curve -Receiver Operating Characteristics is used to visualize theperformance of the multi-classification classifier. This performance is measured atvarious threshold settings.ROC is a probability curve and AUC represents the degree or measurement of separability. The higher the AUC, the higher the correct prediction.

### B. Accuracy:

One parameter for assessing classification models is accuracy. Informally, accuracy is the fraction of predictions of the model got right. Accuracy = Number of correct predictions/Total number of predictions. For binary classification, accuracy can also be calculated in terms of positives and negatives as follows in Eq (1):

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN= False Negatives. The higher the accuracy, the performance of the model is high, meaning, the model is able to classify correctly the output.

## 6. LIMITATIONS

The available systems have a number of drawbacks, including

Garcia [3]et al., created model cannot be used after the first 24 h of observation and data obtained from a single institute of MIMIC-III. Davide et al., author's model is efficient in locating the admissions of live patients, but it is ineffective in locating the admissions of dead patients. One more disadvantage is the absence of a period between a septic episode and the disease. Ran Lix et al Here clinical state transitionis studied has limitations like single institute data and outside ICU, no sufficient datato evaluate Sepsis-3 criteria.

Mostafa et al., the model needs independent validation with laboratory experiments required to confirm the insights. Zma et al ., used multi-center studies, the vital variations within the septic patients, and its evaluation is very much required. The majority of studies had disadvantages like small data set size, retrospective analysis, validation in prospective settings, and single institute database.

## 7. CONCLUSION

A detailed review is conducted to understand the work done in the area of applying machine learning and deep learning concepts in the early prediction of sepsis of different types in different age groups. The selection of data from different sources and the selection of features from the selected data sources differ from one research work to another. Individual paperwork advantages and limitations are discussed in detail. Such as traditional approaches lack sensitivity and have a fine line between infection and inflammatory conditions. Hence ML and DL methods are proposed to automate the task of identifying sepsis at different stages.

## REFERENCES

[1] Moor, M., Rieck, B., Horn, M., Jutzeler, C.R., Borgwardt, K.: Early prediction of sepsis in the icu using machine learning: a systematic review. Frontiers in medicine **8**, 607952 (2021)

[2] Bunn, C., Kulshrestha, S., Boyda, J., Balasubramanian, N., Birch, S., Karabayir,I., Baker, M., Luchette, F., Modave, F., Akbilgic, O.: Application of machine learning to the prediction of postoperative sepsis after appendectomy. Surgery **169**(3), 671–677 (2021)

[3] García-Gallo, J.E., Fonseca-Ruiz, N., Celi, L., Duitama-Muñoz, J.: A machine learning-based model for 1-year mortality prediction in patients admitted to an intensive care unit with a diagnosis of sepsis. Medicina intensiva **44**(3), 160–170 (2020)

[4] Chicco, D., Jurman, G.: Survival prediction of patients with sepsis from age, sex, and septic episode number alone. Scientific reports **10**(1), 1–12 (2020)

[5] Rodriguez, A., Mendoza, D., Ascuntar, J., Jaimes, F.: Supervised classification techniques for prediction of mortality in adult patients with sepsis. The American Journal of Emergency Medicine **45**, 392–397 (2021)

[6] Fujii, T., Luethi, N., Young, P.J., Frei, D.R., Eastwood, G.M., French, C.J., Deane, A.M., Shehabi, Y., Hajjar, L.A., Oliveira, G., *et al.*: Effect of vitamin c, hydrocortisone, and thiamine vs

hydrocortisone alone on time alive and free of vasopressor support among patients with septic shock: the vitamins randomized clinical trial. Jama **323**(5), 423–431 (2020)
[7] Rudd, K.E., Johnson, S.C., Agesa, K.M., Shackelford, K.A., Tsoi, D., Kievlan, D.R., Colombara, D.V., Ikuta, K.S., Kissoon, N., Finfer, S., *et al.*: Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the global burden of disease study. The Lancet **395**(10219), 200–211 (2020)
[8] Adams, R., Henry, K.E., Sridharan, A., Soleimani, H., Zhan, A., Rawat, N., Johnson, L., Hager, D.N., Cosgrove, S.E., Markowski, A., *et al.*: Prospective, multi-site study of patient outcomes after implementation of the trews machine learning-based early warning system for sepsis. Nature medicine **28**(7), 1455–1460(2022)
[9] Hansen, C.L., Viboud, C., Chaves, S.S.: The use of death certificate data to characterize mortality associated with respiratory syncytial virus, unspecified
bronchiolitis, and influenza in the united states, 1999–2018. The Journal of Infectious Diseases **226**(Supplement 2), 255–266 (2022)
[10] Schinkel, M., Paranjape, K., Panday, R.N., Skyttberg, N., Nanayakkara, P.W.: Clinical applications of artificial intelligence in sepsis: a narrative review. Com- puters in biology and medicine **115**, 103488 (2019)
[11] Posthuma, L., Downey, C., Visscher, M., Ghazali, D., Joshi, M., Ashrafian, H., Khan, S., Darzi, A., Goldstone, J., Preckel, B.: Remote wireless vital signs mon- itoring on the ward for early detection of deteriorating patients: a case series. International journal of nursing studies **104**, 103515 (2020)
[12] Rhee, C., Chiotos, K., Cosgrove, S.E., Heil, E.L., Kadri, S.S., Kalil, A.C., Gilbert, D.N., Masur, H., Septimus, E.J., Sweeney, D.A., *et al.*: Infectious diseases society of america position paper: recommended revisions to the national severe sepsis and septic shock early management bundle (sep-1) sepsis quality measure. ClinicalInfectious Diseases **72**(4), 541–552 (2021)
[13] Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., Calvert, J., Das, R.: Evaluation of a machine learning algorithm for up to 48- hour advance prediction of sepsis using six vital signs. Computers in biology andmedicine **109**, 79–84 (2019)
[14] Zhao, T., Zheng, Y., Wu, Z.: Feature selection-based machine learning modeling for distributed model predictive control of nonlinear processes. Computers & Chemical Engineering **169**, 108074 (2023)
[15] Biau, G.: Analysis of a random forests model. The Journal of Machine Learning Research **13**(1), 1063–1095 (2012)
[16] Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., Xiang, Y.: Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid sub- tropical climates: A case study in china. Energy conversion and management **164**, 102–111 (2018)
[17] Pisner, D.A., Schnyer, D.M.: Support vector machine. In: Machine Learning, pp. 101 121 (2020)