

Outcome Prediction of Patients for Different Stages of Sepsis Using Machine Learning Models



Pankaj Chaudhary, Deepak Kumar Gupta, and Samayveer Singh

Abstract Sepsis is a major challenge in the field of medical science. It affects over a million patients annually and also increases the mortality rate. Generally, sepsis condition is not identified easily. Thus, an intensive analysis of patients is required for identifying sepsis in the Intensive Care Unit (ICU). In this research work, an outcome prediction based machine learning models for identifying different stages of sepsis is proposed. Machine Learning (ML) models can help to predict the current stage of sepsis using existing clinical measurements like clinical laboratory test values and crucial signs in which patients are at high risk. We explore four ML models namely XGBoost, Random Forest, Logistic Regression, and Support Vector Machine by utilizing clinical laboratory values and vital signs. The performance evaluation of the proposed and existing techniques is performed by considering the same dataset. These models achieve an AUC (Area under the Curve) 0.95, 0.91, 0.76, and 0.93, respectively, for recognition of sepsis. Experimental results demonstrate that the XGBoost model with 10-fold cross-validation performs well than other models across all the performance metrics.

Keywords Sepsis · Machine learning · Medical science · Intensive care unit

1 Introduction

Sepsis is a fatal medical condition that usually occurs in patients of the ICU. It occurs when releases chemicals in human blood are not able to tackle the infection [1], and is one of the crucial causes of death in the ICU. As shown by the studies the

P. Chaudhary (✉) · D. K. Gupta · S. Singh
Dr. B.R, Ambedkar National Institute of Technology Jalandhar, Jalandhar, Punjab, India
e-mail: pankajchy037@gmail.com

D. K. Gupta
e-mail: guptadk@nitj.ac.in

S. Singh
e-mail: samays@nitj.ac.in

probability of occurrence is rising with every passing year. Sepsis caused an uncontrolled response of the immune system to tackle inflammations and its affects are tissue damage, organ dysfunction, and eventually death [2]. Recovery from sepsis is possible, but this needs careful selection of essential medicines and the timely administration of medication. It has already become a significant international health burden because of higher treatment costs, as well as hospitalization expenses. However, precise recognition of the risk associated with sepsis, and the selection of suitable antibiotics would play a major role in the depletion of mortality rate and cost among the patients in an ICU. Currently, existing screening methods such as the modified early warning system (MEWS) [3], systemic inflammatory response syndrome (SIRS) [4], sequential organ failure assessment score (SOFA score), etc., are not appropriate to recognize sepsis patients [5, 6]. Without timely treatment, sepsis can increase to the next level called septic shock, which has a hospital mortality rate greater than 40% [7]. Even though the death rate has decreased in recent years, the incidence of sepsis is increased rapidly in ICU, so that overall deaths are increasing. Sepsis is preventable if treated early. Identify sepsis as quickly as possible by using the data from past patients, it avoids the need for tools to have to experiment on new patients. Generally, electronic health records (EHR) data is considered for every measurement in a lab test and inform the patients if required. Using various ML models to analyze data of thousands of patients to identify important signs and symptoms that appear in patients with sepsis. Also needs to do is to figure out how to think about every signal in the context of every other signal in the human bodies. When a patient's body is septic it affects patient kidney it deteriorates your kidney capability to filter out creatinine, so creatinine level increases in patient's bodies. But there are numerous factors that can affect patient kidney ability to filter out creatinine. For example, if patient have chronic kidney disease patients are very likely to have high creatinine levels. So now ML models are having the ability to recognize the complex patterns within data and use those correlations for classification to figure out is your creatinine high because of sepsis or because of chronic kidney disease or numerous other factors that need to high creatinine levels. Every single signal that exists in the EHRs is important to think about every signal in the context of every other signal to identify signs and symptoms that occur more often in patients with sepsis and those without sepsis. Doctors sometimes find disease very late and that is the difference between life and death. Machine learning model provides doctors a much larger window to come in and intervene in order to prevent organ dysfunction and mortality. In many we might not need new measurements, the signs and symptoms were already present in the database. We use a machine learning model for discovering these signs and symptoms from the available database and learn something's that human eye unable to visualize. ML model learns much faster if it had a lot of patient's data available. Availability of abundant quantities of data improves the performance of the model. Various published epidemiological studies have already divulged the possibility of loss associated with sepsis. Significant number of studies have also confirmed that ML can recognize sepsis using the crucial clinical values [8]. There are many machine learning algorithms for binary classification and prediction applied to predict the most killer diseases like sepsis and its variant stages. In this

work, we employed the eXtreme Gradient Boosting (XGBoost), Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM) ML models to diagnose different stages of sepsis. Using the relevant clinical information that is obtained from sepsis patients during ICU period, we tend to train the ML models to calculate various useful metrics like sensitivity, specificity, and AUC in order to compare the performance of the models. In this study, XGBoost model attained highest AUC for the prediction of sepsis and also show the improvement of AUC over the existing methods in the literature.

The paper is organized as follows: In Sect. 2, we review the relevant existing research. In Sect. 3, we illustrate the experimental setup. In Sect. 4, we presents a machine learning model and proposed work and Sect. 5 discusses the results in comparison to existing works. Finally, Sect. 6 concludes the paper with some future recommendations.

2 Related Works

Use of routinely available physiological, pathology reports, biomedical signals, etc. are used to identify development of sepsis from one stage to another stage in an ICU [9]. Some studies mainly focused on the existing score-based system to identify patients with sepsis. These scores such as the SIRS, MEWS, Acute Physiology and Chronic Health Evaluation (APACHE) [10], Simplified Acute Physiology Score (SAPS), the Multiple Organ Dysfunction Score (MODS) [11], and SOFA [12]. In 2016, Mitchell et al. developed a LR model framework to distinguish non-ICU and ICU ward patients. In this research work, MIMIC-II database is used for the ICU patients. Their applied model attained an AUC of 73.9%, 56.2% for non-ICU, and ICU patients, respectively [13]. The recent research work mainly focused on ML techniques for detection of sepsis and related conditions. Several studies report that machine learning algorithm for sepsis prediction and detection reduced patient mortality and length of stay in an ICU. Desautels et al. proposed ML Insight method for the prediction of sepsis in 2016. MIMIC-III dataset was considered for this research. Desautels et al. compared the performance achieved by the proposed method to other existing manual scoring systems such as SOFA or SIRS. AUC score of 0.88 is achieved by using Insight method. This method is restricted to ICU patients and aged of patients are ≥ 15 years [14]. In 2018, Mao et al. proposed an ensemble boosting method to construct a sepsis prediction model by utilizing only six vital signs that are routinely checked and measured at medical facilities. Utilization of less number of vital signs reduced the cost and complexity of the model. Their model achieved an accuracy of 92% to classify patients with shock and no shock, and an accuracy of 96% achieved to predict the event 4 h before the onset of septic shock. Mao et al. achieved an AUC of 0.96 for the prediction of septic shock and an AUC of 0.85 for the prediction of severe sepsis. Further AUC of 0.92 for detection of sepsis and AUC of 0.87 for detection of severe sepsis [15]. In 2015, Guillen et al. in his research applied various ML models for the identification of early stages of sepsis.

For model training and evaluation purposed MIMIC-II database was used. 24 h' time period were considered for the observation of pathological samples. Samples collected 2 h before the occurrence of severe sepsis are considered as useful samples other than that are discarded. The data outside the 24 h's period are not considered for the model evaluation. Data imputation techniques were used to handle the missing clinical samples during the time period of 24 h. The model evaluation concluded that the SVM model performed better than logistic model trees. The SVM model achieved an AUC of 0.87 for the recognition of different stages of sepsis [16]. In 2018 Wang et al. proposed the fruit fly optimization algorithm that enhance the performance of existing kernel extreme learning machine (KELM) method for diagnosis of sepsis. Random Forest classifier is used for feature selection. The proposed method got 89.57% sensitivity, and 65.77% [17]. In 2018, Nemati et al. proposed a framework to predict sepsis using Artificial Intelligence methods. The proposed method is useful for identifying the patients who are at risk for sepsis and initiating appropriate treatment, prior to any clinical manifestations, would have a significant impact on the reduction of death rate and the cost burden of sepsis patients. The best performance achieved for predicting tSOFA 4 h in advance AUROC of 0.87 and sensitivity of 0.85 which was slightly higher than predicting tSepsis with AUROC of 0.85 [18]. Recently in 2019, Moor et al. demonstrated that Multi-task Gaussian process temporal adapter model to handle a non-uniform spaced time series data in early detection of sepsis [19].

3 Experimental Setup

In this section, we will discuss the dataset exploration, data cleaning, and data extraction and gold standards.

3.1 Dataset Exploration

This work uses the publically available “*Sepsis Skaraborg study*” dataset [20] collected between 2011 and 2012. A total of 1572 sepsis patients (only adults are considered) with an average age of 67.3 years were enrolled in this study (out of which 55.6% of whom were male patients and 44.4% were female patients) from an ICU or emergency ward. The data set includes comprehensive clinical data such as vital signs, appropriate medications, clinical measurements,, description of Sepsis-2 criteria, Sepsis-3 criteria hospital length of stay, survival data etc.

3.2 *Data Cleaning and Data Extraction*

The collected dataset were available in the form of xlsx file format. Dataset contain some categorical features like gender, age group, etc., we first convert categorical features into numeric features, because some model can handle only numeric features. We create dummy features for the transformation of categorical features into numeric features. After features transformation the next step of data processing is handling the missing values. We use data imputation technique to replace the missing measurement with another value by a simple average mean method. Missing data imputation improved the overall model performance. For more data exploration, we used a kernel density estimation [21] for outlier detection. Once the data is processed we split the sepsis dataset into train and test set in the ratio of 80:20 by using python ML library. After splitting process we train the ML model to predict onset sepsis.

3.3 *Gold Standards*

In this analysis, we followed the gold standard and new definitions of sepsis to train and test the dataset. ML model performance was tested on training and testing set according to various gold standards. SIRS (Systemic Inflammatory Response Syndrome), and SOFA (Sequential [Sepsis-related] Organ Failure Assessment) at least one of these gold standards satisfied to meet the definition of sepsis [22, 23]. In fact, we were concerned about the whole time series data of each patient's entry to observe the ICU patients.

Following features are considered for applying the ML techniques:

- 6 Vital signs: Respiratory Rate, Oxygen Saturation, Systolic Blood Pressure, Heart Rate, Temperature, Hemoglobin.
- 10 Lab test results: Leukocyte, Procalcitonin, Neutrophil-lymphocyte count ratio, White Blood Cell (WBC), C-reactive protein, Creatinine, Positive blood culture, P-lactate, Systemic inflammatory response syndrome, SOFA score.
- Location type: Emergency department or ICU.
- Progressive stages label preceding to septic shock: infection, inflammation, organ dysfunction, Bacterial sepsis, severe Bacterial sepsis/septic shock based on Sepsis-2 definition, and Bacterial sepsis based on new Sepsis-3 definition [24]: all are identified by machine learning experts system.

4 **Machine Learning Models Design**

In this section, we explore four binary classification model for performance comparison namely XGBoost, Random Forest, Logistic Regression, and SVM.

4.1 XGBoost (EXtreme GRAdient Boosting)

This study investigates the effectiveness of XGBoost model for the identification of sepsis. It is more efficient than the gradient boosting method. Boosting is an ensemble method in which every new model are added to correct the errors made by the previous models [25, 26]. It supports both the linear model solver and tree learning algorithms. XGBoost is faster than the other existing boosting implementations because it allows parallel computation on single machine. The main motivation of using XGBoost is the ability to fine-tune hyper-parameters and Built-in Cross-Validation to improve the performance of the model to predict the sepsis. Workflow analysis of the proposed method is demonstrated in Fig. 1. For cross-validation we select $k = 10$ and whole dataset is split into stratified bins of equal size, which was used to train and evaluate our models. Python Scikit-Learn library and XGBoost package [27] were used for the implementation of model.

Model evaluation. To use the XGBoost to classify sepsis disease, the minimum information we need to provide to the model are Input features, Target variable, Objective function, Number of iteration (Number of tree added to the model for classification), `n_estimators`, `learning_rate`, and `max_depth` of the tree. We try to change the value of parameters in order to achieve the optimal performance of the model. To train the model we split the data set into the ratio of 80:20, respectively,

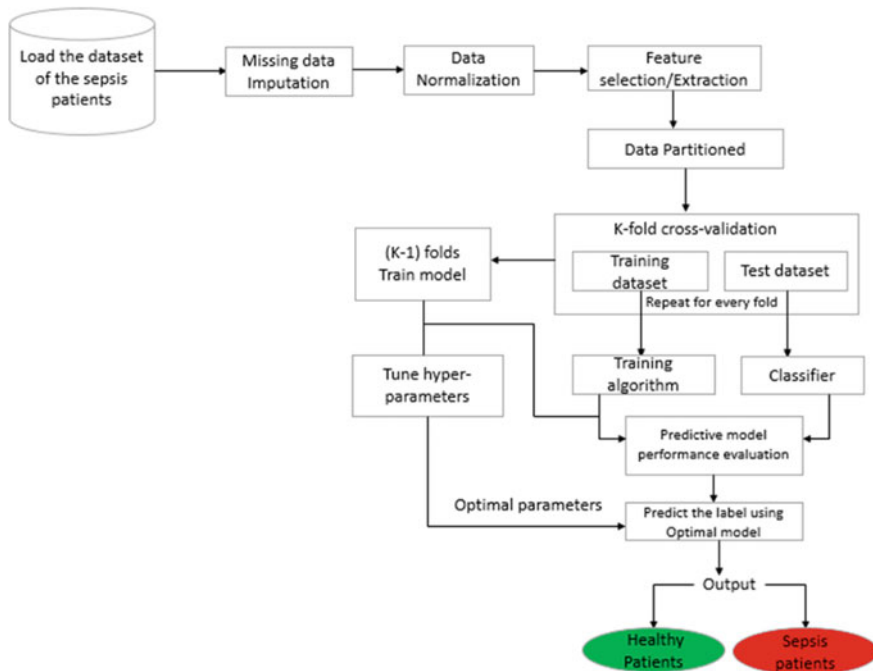


Fig. 1 Experimental workflow of model

and then applied 10-fold cross-validation and GridSearchCV to select the different parameter values to optimize the performance of the model.

Model specification. Distribution of train and test data set. To train the model, we need to optimize the function to get low test error. For optimization typically, we use: Log Loss for binary classification ($L1$), and mlog loss for multi-classification ($L2$) [27].

$$L1 = -\frac{1}{N} \sum_{i=1}^N (b_i \log(p_i) + (1 - b_i) \log(1 - p_i)) \quad (1)$$

In the above equation, b is denoted as the label and $p(b)$ is a predicted probability.

$$L2 = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M b_{i,j} \log(p_{i,j}) \quad (2)$$

In the above equation, N is the total number of samples and M is the total number of outcomes that are possible for a given situation.

4.2 Random Forest

This section outlines the details of a Random Forest (RF) [28] model. It is an ensemble ML method. With different initialization RF consist set of decision trees and trained in parallel manner. RF can be thought of as an extension to the decision tree base classifiers. The Random Forest ensemble classifier [29] has been used on many datasets spanning different environments and industries. The RF ensemble is preferred over other ensembles method because it is simple, can be easily parallelized, is comparatively strong to outliers, and noise, and it is comparatively faster than bagging or boosting techniques. RF Classifier being an ensemble algorithm tends to give a more accurate result because it works on the principle of weak estimators. Once all the weak estimators are merged it forms a strong estimator. Even though one or more decision trees are suffering from noise, the overall result could not be affected. The main reason behind choosing the RF model for prediction of sepsis are included: It runs efficiently on large dataset and model does not suffer from Underfitting. For the implementation of RF we first split our training and testing data set in the ratio of 80:20, respectively. Then, we import RandomForestClassifier from sklearn python library. We took $n_estimators = 20, 100, \text{ and } 200$, respectively, and change the other RF parameters during the model training to check the performance of the model.

4.3 *Logistic Regression*

Logistic Regression (LR) is a supervised machine learning classifier [29, 30]. It is a widely used technique for analyzing laboratory data. LR model is simple and highly interpretable and it is suitable for the baseline model to compare the results with other machine learning models. LR requires the dependent variable (target variable) should be in binary 0 or 1, to identify whether patients suffer from sepsis or not. For building an LR model in python, we import the linear model Logistic Regression package. To train the model we split the data set into the ratio of 80:20, respectively, and then applied 10-fold cross-validation technique and Grid Search CV to select the different parameter values to get the best performance of the model.

4.4 *Support Vector Machine*

This section outlines the implementation of a Support Vector Machine (SVM) [31] model. It is a classification method used to classify a patient into a new group. Our task is to identify patients with sepsis or patient without sepsis using SVM model. This method is suitable for binary classification problems, this means that 0 denote healthy patients and 1 denote patients with prone to sepsis or septic shock. Data of the different patients are similar, so performing linear SVM data separation is not efficient in this case. In this situation SVMs can efficiently perform a non-linear data separation with the use of the SVM kernel trick. Entire data separation problem mapped into multi-dimensional space. To minimize an error SVM generates optimal hyperplane in an iterative fashion. We Scikit-Learn python library for the implementation of SVM. We split train-test data in the ration of 80:20. To boost the performance of the model we applied the Gaussian Kernel method for the implementation of kernel SVM. For classification we select the two important parameters they are: First parameter is C: Inverse of the strength of regularization. As the value of 'C' increases the model gets overfit and the value of 'C' decreases the model get underfit. We select $C = 100$ for SVC (support vector classification) Second important parameter is γ : Gamma (used only for RBF kernel). As the value of ' γ ' increases the model gets overfit and the value of ' γ ' decreases the model underfit.

5 **Experimental Results and Discussions**

This section will introduce the performance analysis of the ML model that is trained for the identification of different sepsis conditions. Further, we will compare the performance of four binary classification models, namely XGBoost, RF, LR, and SVM based on the selected performance metrics.

5.1 Evaluation Metrics

In this section, we will discuss all the useful performance metrics to analyze the performance and predictive power of models. AUC curve (Area under the curve), accuracy, recall, and precision were computed to evaluate the models. Confusion matrix could be a helpful metric to explain the performance of a classification model on a set of test data. Confusion matrix table is divided into four parts such as True Positives (TP): Count total number of true positives values ($y_{\text{true}} = 1, y_{\text{pred}} = 1$) and True Negatives (TN): Compute number of true negatives values ($y_{\text{true}} = 0, y_{\text{pred}} = 0$) are the observations which were correctly predicted. The terms False Positives (FP): Count total number of false positives values ($y_{\text{true}} = 0, y_{\text{pred}} = 1$) and False Negatives (FN): Compute number of false negatives values ($y_{\text{true}} = 1, y_{\text{pred}} = 0$) can lead to confusion.

Recall (Sensitivity). Evaluate the percentage of patients that really had sepsis was properly diagnosed by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

Specificity. Clinical test to correctly identify healthy patients.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (4)$$

AUROC (Area under the Receiver Operating Characteristics). ROC curve is a plot of the true positive rate (TPR) along with Y-axis against the false positive rate (FPR) along with X-axis. ROC indicates the probability curve and AUC quantify the separability. To select the AUC metric for a classification problem we need to know about the possible range of AUC that defines in the interval of [0, 1]. 100% TPR and 0% FPR indicate the perfect classifier, and ROC curve passes through the upper left corner of the square, and AUC values less than 0.5 indicate worthless classifier [34]. AUC is mainly used in the situation when accuracy misleads the classification problem, even we get 99% accuracy that doesn't mean the model is absolutely good for prediction of disease. In that case, AUC is considered good for the classification problem. Golden standard in the field of medicine indicates $\text{FPR} = 0$ and $\text{TPR} = 1$.

5.2 Results Summary

ML model was trained and tested on the sepsis patient's data set. This dataset contains a total of 1572 patients out of which 1257 patients were selected for the training dataset and 315 patients for test dataset. Out of four models the highest AUC achieved is 0.95 and lowest AUC achieved is 0.76, and the average AUC of 0.88 achieved for

the identification of the different levels of sepsis. The AUROC curve of each model is shown in Fig. 2.

XGBoost, RF, LR, and SVM models achieves AUC 0.95, 0.91, 0.76, and 0.93, respectively, for recognition of sepsis. We select linear kernel as well as Gaussian Kernel to build the SVM model for recognition of sepsis. We got an AUC of 0.76 and 0.93 for linear and Gaussian Kernel, respectively. Gaussian Kernel performed better on this dataset as compared to linear kernel for SVM model. We improved the prediction of sepsis by using ML models. Table 1 shows the results of different existing machine learning techniques in terms of their AUC for the identification of sepsis. The AUC value for the Desautels et al. (2016) [14], Guillen et al. (2015) [16], Balduino et al. [32], and Masino et al. (2019) [33] is 0.74, 0.87, 0.85, and 0.82 using Insight: 4 h, SVM, XGBoost, and RF, respectively (Fig. 3).

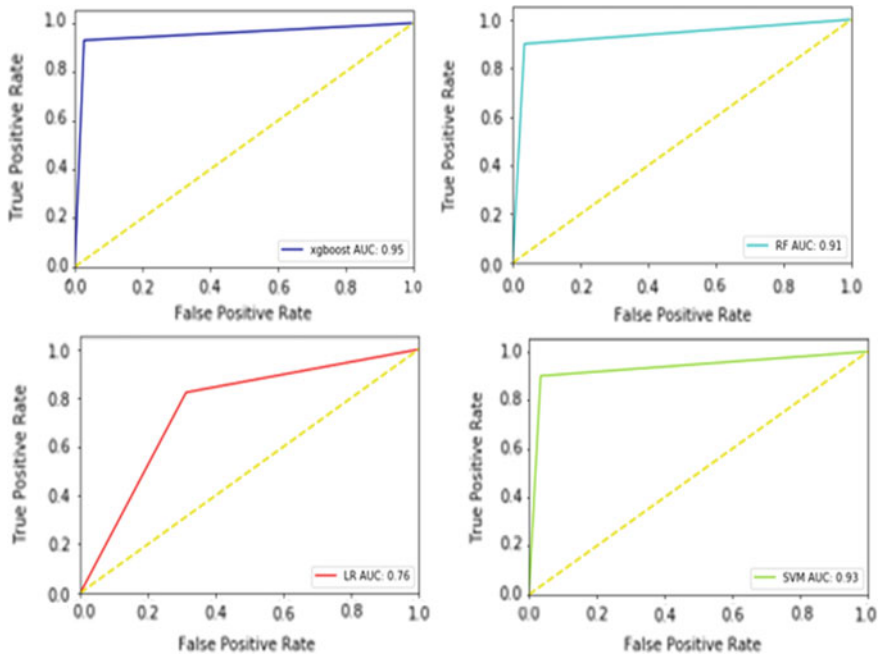


Fig. 2 ROC curve of XGBoost, RF, LR, and SVM

Table 1 Evaluation of performance of the model in terms of sensitivity, specificity, and AUC

Model	Sensitivity	Specificity	AUC
XGBoost	0.92	0.97	0.95
RF	0.88	0.94	0.91
LR	0.82	0.69	0.76
SVM	0.90	0.97	0.93

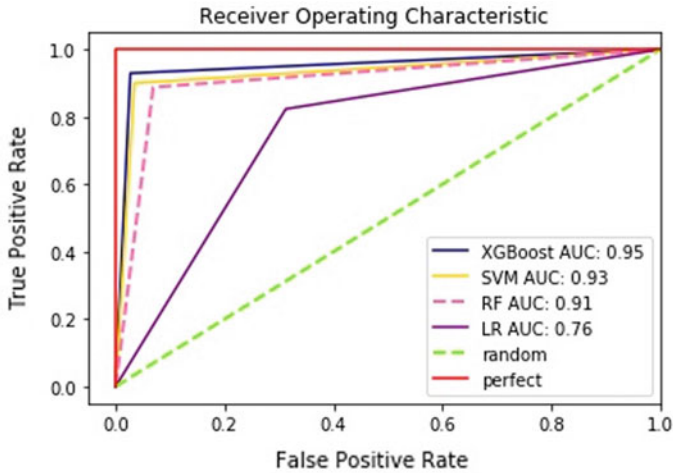


Fig. 3 ROC curve for all the four models for prediction of sepsis

Table 2 Results of the selected studies that applied the ML techniques to identify the sepsis

Paper	Data set	Method	AUC
Desautels et al. (2016) [14]	MIMIC-III	Insight: 4 h	0.74
Guillen et al. (2015) [16]	MIMIC-II	SVM	0.87
Balduino et al. (2018) [32]	Geisinger Healthcare System	XGBoost	0.85
Masino et al. (2019) [33]	Hospital of Philadelphia	RF	0.82

From Table 2, we see that the highest AUC value is 0.87 in [16] that is lower than our proposed work. However, it's tough to make a direct comparison as a dataset, and size of sample were used during this analysis are completely different.

6 Conclusion

In this research work, we have tested sepsis dataset using various ML model for accurate prediction of sepsis. First, we have used the imputation process to deal with missing data to improve the model performance then imputing using matrix factorization yields better predictive accuracy than imputing with simpler approaches like means. After that we used different packages to build models like XGBoost, LR, SVM, and RF. The XGBoost, Random Forest, Logistic Regression, and SVM ensemble model achieve 0.95, 0.91, 0.76, and 0.93 AUCROC for sepsis, respectively.

In summary, the result of XGBoost ensemble model improves the overall performance and delivers a more accurate classification as compared to other involved models in this study for the prediction of sepsis. The models presented in this work show promise in their ability to decrease morbidity and mortality rates resulting from septic shock and improve the outcome for sepsis patients.

The major limitations of this research work are the size of samples, so the work can be extended by using more data samples to improve the prediction of the sepsis. Another issue in this work is to focus on only the ICU in a single medical center and a limited period of time. In the future, researchers can collect more relevant laboratory specimens from multiple medical center databases to establish an error-free sepsis prediction model. We believe this model will be useful in the hospital environment to predict sepsis disease.

References

1. Novosad SA, Sapiano MR, Grigg C et al (2016) Vital signs: epidemiology of sepsis: prevalence of health care factors and opportunities for prevention. *MMWR Morb Mortal Wkly Rep* 65(33):864–869
2. van Wyk F, Khojandi A, Kamaleswaran R (2019) Improving prediction performance using hierarchical analysis of real-time data: a sepsis case study. *IEEE J Biomed Health Inf* 23(3):978–986
3. Subbe CP, Slater A, Menon D, Gemmell L (2006) Validation of physiological scoring systems in the accident and emergency department. *Emerg Med J* 23 (11):841–845
4. Lamping F, Jack T, Rübsamen N et al (2018) Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children—a data-driven approach using machine-learning algorithms. *BMC Pediatrics* 18(1):1471–2431
5. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M et al (2016) The third international consensus definitions for sepsis and septic shock (Sepsis-3). *J Am Med Assoc* 315(8):801–810
6. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A et al (2016) Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *J Am Med Assoc* 315(8):762–774
7. Ho JC, Lee CH, Ghosh J (2014) Septic shock prediction for patients with missing data. In: *ACM transactions on management information systems (TMIS)* 5(1):1–1
8. Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, Angus DC, Reinhart K (2016) Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am J Respir Crit Care Med* 193:259–272
9. Faisal M, Scally A, Richardson D, Beatson K, Howes R, Speed K, Mohammed MA (2018) Development and external validation of an automated computer aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Crit Care Med* 46(4):612–618
10. Sadaka F, EthmaneAbouElMaali C, Cytron MA, Fowler K et al (2017) Predicting mortality of patients with sepsis: a comparison of apache ii and apache iii scoring systems. *J Clin Med Res* 9(11):907–910
11. Badrinath K, Shekhar M, Sreelakshmi M, Srinivasan M et al (2018) Comparison of various severity assessment scoring systems in patients with sepsis in a tertiary care teaching hospital. *Indian J Crit Care Med* 22(12):842–845
12. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H et al (1996) The SOFA (Sepsis-related organ failure assessment) score to describe organ dysfunction/failure on

- behalf of the working group on sepsis-related problems of the European society of intensive care medicine. *Intensive Care Med* 22(7):707–710
13. Mitchell S, Schinkel K, Song Y, Wang Y (2016) Optimization of sepsis risk assessment for ward patients. In: 2016 IEEE systems and information engineering design symposium, pp 107–112. IEEE, Charlottesville
 14. Desautels T, Calvert J, Hoffman J et al (2016) Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. *JMIR Med Inf* 4(3). <https://doi.org/10.2196/medinform.5909>
 15. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, Shieh L, Chettipally U, Fletcher G, Kerem Y et al (2018) Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and icu. *BMJopen* 8(1)
 16. Guillen J, Liu J et al (2015) Predictive models for severe sepsis in adult icu patients. *IEEE Systems and Information Engineering Design Symposium*, IEEE
 17. Wang X, Wang Z, Weng J, Wen C, Chen C, Wang X (2018) A new effective machine learning framework for sepsis diagnosis. *IEEE Access* 6:48300–48310
 18. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG (2018) An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med*, 46 (4), 547–553 (2018)
 19. Moor M, Horn M et al (2019) Early recognition of sepsis with Gaussian process temporal convolutional networks and dynamic time warping. In: *Proceedings of machine learning research XX*, pp 1–IX
 20. Sepsis biomarker study Homepage, <https://doi.org/10.5281/zenodo.823967>. Last accessed 7 Jan 2019
 21. Latecki LJ, Lazarevic A, Pokrajac D (2007) Outlier detection with kernel density functions. In: *Conference MLDM 2007, LNCS, Springer, Leipzig*, vol 4571, pp 61–75
 22. Fullerton JN, Thompson K, Shetty A, Iredell JR, Lander R, Myburgh JA, Finfer S (2017) New sepsis definition changes incidence of sepsis in the intensive care unit. *Crit Care Resuscitation* 19(1):9–13
 23. Barton C, Chettipally U, Zhou Y, Jiang Z, Lynn-Palevsky A et al (2019) Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Comput Biol Med* 109:79–84
 24. Shankar Hari M, Harrison DA, Rubinfeld GD, Rowan K (2017) Epidemiology of sepsis and septic shock in critical care units: comparison between sepsis-2 and sepsis-3 populations using a national critical care database. *British J Anaesth* 119:626–636. <https://doi.org/10.1093/bja/aex234>
 25. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. *KDD 2016, ACM, San Francisco*
 26. Freund Y, Schapire RE (1999) A short introduction to boosting. *J Japanese Soc Artif Intell* 14(5):771–780
 27. XGBoost Python package homepage. <https://xgboost.readthedocs.io>. Last accessed 21 July 2019
 28. Breiman L (2001) Random forests. *Mach Learn* 45 (1):5–32
 29. Mitchell T (1997) *Machine learning*. McGraw-Hill
 30. Andrew Y, Jordan Ng (2002) On discriminative versus generative classifiers: a comparison of logistic regression and naive Bayes. In: *Dietterich TG, Becker S, Ghahramani Z (eds) Advances in NIPS 14, MIT Press, Cambridge*, pp 841–848
 31. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: *Proceedings of ECML-98, 10th European conference on machine learning*, Springer, Dortmund, pp 137–142. <https://doi.org/10.1007/bfb0026683>
 32. Balduino R, Bogle B, Hazard A (2019) Using machine learning to predict outcomes for sepsis patients: <https://medium.com/inside-machine-learning/using-machine-learning-to-predict-outcomes-for-sepsis-patients-cde5625ce9cb>. Last accessed 21 May 2019

33. Masino AJ, Harris MC, Forsyth D (2019) Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data. In PLoS ONE 14(2)
34. Cali C, Longobardi M (2015) Some mathematical properties of the ROC curve and their applications. *Ricerche mat* 64(2):391–402