Using Gated Recurrent Units Models for Early Prediction of Sepsis in The Intensive Care Unit

Xuze Zhao

Tianjin University

Bo Qu (✉ doctor_qb@tju.edu.cn)

Tianjin University

Research Article

Keywords: sepsis, deep learning, machine learning, artificial intelligence, gated recurrent units

Posted Date: March 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-259370/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Using Gated Recurrent Units Models for Early Prediction of Sepsis in the Intensive Care Unit

Xuze Zhao,¹ and Bo Qu²

¹Tianjin University; Tianjin Key Laboratory for Disaster Medicine Technology, Tianjin China.

²Tianjin University; Tianjin Key Laboratory for Disaster Medicine Technology, Tianjin China.

Correspondence should be addressed to Bo Qu; doctor_qb@tju.edu.cn

Abstract

Background: Sepsis is one of the major causes of mortality in hospitalized patients. Therefore, a reliable means of predicting sepsis onset is of great importance. The purpose of this study was to develop a gated recurrent unit (GRU) based model and explore whether it can improve model performance in predicting sepsis for up to 6h from the time of admission in intensive care units (ICU) compared to traditional sepsis prediction methods.

Methods: The data used for model development in this study were from retrospective MIMIC-III dataset, restricted to intensive care units (ICUs) patients aged between 15 and 89. Model performance of GRU model were compared to logistic regression (LR), support vector machine (SVM), random forest (RF), and extreme gradient boost (XGBoost). The area under the receiver operating characteristic (AUROC) measures the performance of the prediction capability of the models.

Results: A total of 31297 MIMIC-III cases are included in this article and 4008 cases had encountered sepsis while 27289 had not. As for the AUC (0.801 (95% CI):

0.760-0.841), 0.782 (95%CI: 0.743-0.821), 0.775 (95%CI: 0.736-0.813), 0.771 (95%CI: 0.732-0.809), 0.749 (95%CI: 0.711-0.886),) results of the models, GRU performed best in predicting sepsis.

Conclusions: The present study concluded that by using GRU deep learning method, a more accurate prediction model can be established. The GRU deep learning method we build can prove clinically helpful and assist physicians in tailoring accurate management and treatment for patients with sepsis.

Key Words: sepsis, deep learning, machine learning, artificial intelligence, gated recurrent units

Introduction

Millions of patients are admitted to intensive care units (ICUs) around the world every year. The health information of patients including vital signs, laboratory test results and demographic details are recorded in ICUs to support medical personnel make life-saving proposals[1]. However, high levels of indeterminacy and severe time restriction make the decisions derived from tremendous volume of complicated clinical data tend to be inaccurate. The clinical data can be integrated and interpreted by the artificial intelligence algorithms which can deal with repetitious patient evaluations in real-time far more effectively thus improve timely and targeted diagnosis[2].

The early prediction of sepsis remains challenging in the ICUs settings. Sepsis is a long-standing and life-threatening issue that emerges due to the patients' response to infection, which can lead to tissue damage, organ dysfunction, and even death[3]. It happens when the immune system of patients releases chemicals into the bloodstream system to cope with infection and leads to inflammation all over the patients' body. To this day, sepsis presents high morbidity and mortality and is the major cause of mortality

in the USA, besides, it's the most expensive circumstance associated with in-hospital stay, which accounts for nearly \$24 billion annually[4]. During the second decade in the 21st century, the sepsis incidence worldwide has reached 0.43% and approximately 6 million deaths every year were caused by the development of sepsis. Furthermore, 4.2 out of 30 sepsis cases occurred in newborns and kids annually[5].

It's a crucial section to make an early prediction and antibiotic treatment for improving sepsis consequences such that a few moments of delayed intervention led to a considerable rise in mortality. However, the prognostic of sepsis in its early stages is tough for medical personnel due to the heterogeneous nature of infectious insults and the diversity of host responses. In 2016, the whole definition of sepsis was renewed as The Third International Consensus Definitions for Sepsis [6](Sepsis-3) to clarify the state of sepsis and therefore to boost earlier prediction of sepsis, however, the previous effort cannot clean the hurdle of early prediction and treatment of sepsis.

Previous researchers have defined amount of disease scoring systems and diagnostic criteria to detect the emergence of patients in the hospital. One of the early proposed sepsis diagnostic criteria is Systemic Inflammatory Response Syndrome[6](SIRS)(Figure 1), SIRS is confirmed when at least two of the four symptoms in Figure1 are present. Sepsis-3 is a redefinition of sepsis that has been presented lately which posed importance on the priority of the non-hemostatic host response to infection, the potential mortality tremendously over a normal infection, and the urgent demand for the prognostic. A scoring system called Sequential Organ Failure Assessment [7](SOFA) was proposed as a screening mechanism in 1996 which aimed to record the patients' condition in the ICUs and calculate the organ function based on Sepsis-3(Table1). Several medical indexes are utilized in SOFA to conduct the prediction: the mean arterial pressure, serum glucose, bilirubin, PaO₂/FiO₂ ratio,

platelets, and creatinine. If more than two points in the SOFA changes acutely, the consequence of organ failure can be concluded due to the infection throughout the patient's body. The severe SOFA score can result in a rise in the probability of death. These methods utilize tabulation of vast patient vital signs, demographics, and laboratory test results to calculate risk scores. However, most of these score systems were proposed a long time ago. The outstanding performance of them is based on the population or medical level at that time, which means, the change of the medical settings leads to a decline of its ability of prediction. More than that, some researchers have found that when facing the risk of in-hospital death of sepsis cases, these score systems perform badly and can lead to misdiagnosis, even death[8].

In considering the poor performance of the precious score systems, some other models have been presented for predicting the risk of in-hospital death amongst ICU patients with sepsis. Machine learning (ML) methods are widely used in the early prediction of the risk of sepsis. In 2018, Shamim Nemati[9] presented an interpretable machine learning model for accurate prediction of sepsis in the ICU, a total of 65 variables forms the electronic medical record, and real-time clinical data were collected and they were used as input features to a Weibull-Cox proportional hazard model. The corresponding area under the receiver operating characteristic curves (AUROC) reported by the model are higher than 0.79. Calvert[10] developed a high-performance early sepsis prediction method for the general patient population called InSight. InSight is a machine learning-based workflow for sepsis prognostic which computes, in real-time, the risk that a patient will develop sepsis. InSight uses vital signs, age, Glasgow Coma Score (GCS), and pulse oximetry as input variables. This machine learning-based method presented a AUROC of 0.74 in predicting sepsis 4 h before onset. The InSight algorithm demonstrated a better performance than expert scoring in predicting sepsis at

hospitals. Khoshnevisan proposed a recent temporal pattern in conjunction with SVM to construct a robust and interpretable model for the early prediction of sepsis. The model conducts the prediction in two different forms: visit-level (left-align) early diagnosis and event-level (right-align) early prediction. It integrated electronic health records (EHRs), clinical expertise, and various ML algorithms to construct a classifier and the AUROC early prediction performance of this classifier is above 0.85 at all times. Recently, Christopher[11] proposed a temporal convolutional network (TCN) to make early detection of sepsis. In TCN, calculations are processed in a way that each time-step is renewed concurrently. It concluded that the mean ROC of their proposed model has a true positive rate of about 1.0 and a false positive rate of about 0.0, the mean PRC of their model has a precision value of about 1.0 and recall value less than 1.0, which means the model can be regarded as a better choice compared to the models before. Al-Mualemi[12] designed a deep learning base sepsis estimation scheme in 2020. The core architecture of the deep learning network is CNN, and the activation function used for the folding process of the data sequence is rectified linear unit. The outcome from the intelligent section in it can assist the physician to evaluate the patients' condition through certain criteria, which means this scheme reports clinicians' view of the prediction for early detection of sepsis.

Various existing limitations can be found in the aforementioned prediction models. Firstly, most of the previous researches does not account for the temporal developments and real-time data assessment of the sepsis and rely on conventional modeling schemes, such as support vector machine and artificial neural network. Secondly, the latest released definition Sepsis-3, the criterion introduced by Singer, should be utilized instead of older gold standards such as SIRS. Finally, the development of sepsis in a long period is subtle that previous works can hardly learn about the discriminative

patterns of sepsis and analyze the informative development during a patient's stay, for example, only a few changes can be noticed on white cell volume and the temperature of patients at the early stage of sepsis.

To clear these limitations, we present a brand-new method for real-time early prediction of sepsis onset for patients admitted to ICUs

Our main contributions are summarized as follows.

- i) To the best knowledge of our knowledge, this article is the first attempt to use a Gated Recurrent Unit method to conduct prediction to a progressive – sepsis and make comparisons between it against other traditional ML methods.
- ii) We run our models with a vast of variables as input which can achieve a more accurate prediction performance such as vital signs, demographics, and laboratory results while most prior research mainly focused on part of these features.
- iii) We apply the newest definition of sepsis, Sepsis-3, as our gold standard for our predictive algorithm which makes the prediction more valuable for the clinicians and patients.

Methods

Database

This study used the Medical Information Mart for Intensive Care III [13] database version 1.4 (MIMIC III v1.4) as the dataset. MIMIC-III is the only freely accessible critical care database of its kind and it was ratified by the Institutional Review Boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology (MIT). comprises 61,532 intensive care unit stays: 53,432 stays for adult patients and 8,100 for neonatal patients. The data ranges from June 2001 to October 2012. MIMIC-

III includes labeled cases such as vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, and survival data. Every single case in MIMIC-III was provided by medical personnel and data analysts and there was no need to get permission from the patient because of its unidentified health information. Every researcher has to finish an exam to register for the MIMIC-III and be qualified by the officials of MIMIC-III in 14 days after the exam. The website of the National Institutes of Health (NIH), the official site of MIMIC III, has credentialed our study to free access to the database after finishing the exam “Protecting Human Research Participants”, the Record ID of the Completion Report is 40043867.

Gold Standard

The gold standard used in our study is the sepsis definition promulgated by Singer et al in 2016[6]. This study defined sepsis as “life-threatening organ dysfunction caused by a dysregulated host response to infection signified by an acute change in total SOFA score >2 points consequent to the infection.” As for the MIMIC-III database, we utilized the International Classification of Diseases, ninth revision, Clinical Modification (ICD-9)[14] diagnosis codes to label the patients who had been infected. However, ICD-9 is only recorded for a limited amount of complications and tremendously different diseases can often share the same code once they have the same cost. Hence, ICD-9 may present false positives of septic and can hardly demonstrate the real condition of patients. As a result, it has been widely argued that ICD-9 codes cannot be used for determining worthy gold standards for various diseases. Other than that, the SOFA score system was regarded as another criterion to label the patients with sepsis from the infected group.

Inclusion Criteria

The criteria shown in Figure were a step to construct a plausibility filter to the MIMIC-III dataset. Firstly, we set a threshold on the age that the patients aged ranging from 18 to 89 can be included. Next, the patient's stay in ICU should be more than 12 hours to ensure the value of the data. Finally, in the MIMIC-III dataset, there exists missing data for some reasons, to deal with this problem, the variables with data missing rate higher than 20% should be excluded.

Data Collection and Data Preprocessing

As for the feature extraction, this study selected a total of 46 variables as the input features for the sepsis early prediction models, including demographics, laboratory data, vital signs, and others (Table 2). We also summarize a variety of features from the MIMIC-III dataset to calculate some scores for prediction such as SOFA. All the data are extracted from the MIMIC-III dataset using Python3.7 (Anaconda Inc.; Austin, Texas USA) and its packages. For the missing data, we conducted a “carry-forward” method that the most recent value is carried forward to fill the empty place.

Prediction Models

LR

LR is a traditional statistical method utilized to model the probability of a certain category[15]. Mathematically, a logistic regression model owns a dependent feature with two different potential values, where can be labeled as “0” and “1”. The relationship between the predictor variables and the logit of the event can be expressed as the following equation (where ℓ is the logit, b is the base of the logarithm, and β_i are variables of the models):

$$\ell = \log_b \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (1)$$

and the odds can be recovered by exponentiating the logit:

$$\frac{p}{1-p} = b^{\beta_0 + \beta_1 x_1 + \beta_2 x_2} \quad (2)$$

SVM

The support vector machine is a kind of supervised learning model using associated learning algorithms to analyze variables from the dataset for classification and regression analysis[16]. Other than performing linear classification, SVM can effectively deal with a non-linear classification with the so-called kernel trick[17, 18], slightly mapping its input features into high-dimensional feature spaces. SVM can deal with the unlabeled data as well. It's able to find out natural clustering of the data to categorize and map the data to distinct groups.

In this study, to deal with a non-linear classification, we utilized kernel trick of the SVM and we assume a kernel function κ which satisfies $k(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$.

RF

Random forest is an ensemble learning method applied to conduct classification, regression, and other tasks[19, 20]. RF constructs amounts of decision trees at the training period, which is a weak classifier of machine learning methods and outputs the category result of the classification or prediction of the individual decision trees. The random forest can correct for decision trees' drawback of overfitting which may pose an impact on the accuracy of the outcome, so it has been widely used in medical prediction and diagnostic decision making. As for tree growing, it uses the Gini index as the baseline to select proper features. Moreover, RF is a classic black-box machine learning model[21] due to its ability to generate accurate results with a huge range of variables and its little need for intervention in packages like scikit-learn.

XGBoost

Extreme Gradient Boosting (XGBoost) is a kind of ML method which possesses

the effective capability of dealing with missing variables and merging weak prediction algorithms to build a strong one[22]. Since it was invented XGBoost has been widely utilized as the benchmark algorithm in amounts of ML and data mining competitions. For example, in Knowledge Discovery in Database Cup 2015, all of the top-10 winning teams selected XGBoost as their algorithm and the winning solution had been published at Kaggle's blog.

GRU

Recurrent neural network (RNN) is a kind of Artificial Neural Network (ANN) of which connections between nodes can be formed as a directed graph along a temporal sequence[23, 24]. GRU is a gating modification in RNN that was proposed in 2014 by Kyunghyun Cho et al. The general structure of GRU is familiar with long short-term memory (LSTM) except for an additional forget gate to avoid vanishing gradient and exploring gradient. GRU has proved its capability of prediction on certain domains and datasets. The structure diagram of the GRU model we used in this article is shown in Figure 3, the mathematical formulation of Figure is as follows (where x_t is input vector, h_t is output vector, $\hat{h}_t$ is candidate activation vector, z_t is update gate vector, r_t is reset gate vector, and W , U , and b are parameter matrices vector):

$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z) \quad (1)$$

$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r) \quad (2)$$

$$\hat{h}_t = \phi_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \quad (4)$$

The model construction software we used in this article is Python 3.7(Anaconda Inc.; Austin, Texas USA) and its packages, we propose a hypothesis that the GRU algorithm performs better than other ML methods in predicting sepsis.

Statistical Analysis

All the cases involved in this study were split into two groups depending on whether it meets the criteria of sepsis. All models used the `train_test_split` function in Python 3.7(Anaconda Inc.; Austin, Texas USA) `sklearn.model_selection` library to divide the data set into a training set and a test set, the test ratio was 0.3 (7:3), and the random seed was set to 5[25], which means the patients are randomly divided into the training set for training prediction models, and testing set for testing the performance of models. Continuous variables are depicted as median value and interquartile range (25%-75%), categorical variables are depicted as numbers or percentages.

The performance of the ML models in predicting the cases was evaluated based on the confusion matrix of the process. Some of the metrics were measured such as error rate, sensitivity, precision, specificity. Besides, the receiver operating characteristic (ROC) was calculated along with the area under the receiver operating characteristic curve (AUROC). AUROC is an overall metric of distinction of a trained model regarding sensitivity and specificity, and it possesses the capacity to distinguish the cases with sepsis from the others. The metrics used to evaluate the performance are given as follows,

$$Error\ Rate = 1 - \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

TP, TN, FP, FN indicates the true positive, true negative, false positive, and false negative respectively.

All the models were constructed and in Python3.7(Anaconda Inc.; Austin, Texas USA) and the data analysis was conducted in SPSS version 22.0 (IBM Corporation, Armonk, NY, USA).

Results

Baseline Characteristics

After excluding the cases that didn't meet the determined criteria of sepsis or miss clinical data, a total of 31297 HS patients are included in this study, 4008 cases had encountered sepsis and 27289 cases had not respectively. Figure 2 is a flow chart that depicts the procedure for case selection. The average age of the cases with sepsis and the cases without sepsis were 61.12 ± 11.24 and 59.11 ± 13.91 years old respectively. 2052 cases with sepsis are male, which accounted for 51.2% of all the cases, 14081 cases without sepsis are male, which accounted for 48.8% of all the cases without sepsis. Thus, it can be seen that age is a relatively significant variable for sepsis that compared with the healthy people the people with sepsis may be several years older, in contrast, gender is not statistically significant in this study. Table 3 is a summary concluding the comparisons of the statistic results between the patients with and without sepsis from MIMIC-III database.

Model Comparisons

Table 4 shows the AUC, specificity, recall, precision, and error rate according to each model in this research. Figure 4 shows all the confusion matrix of all the models. The average AUROC values of the GRU, SVM, RF, LR, and XGBoost models (Figure 5) were 0.801 (95% confidence interval (CI): 0.760-0.841), 0.782 (95%CI: 0.743-0.821), 0.775 (95%CI: 0.736-0.813), 0.771 (95%CI: 0.732-0.809), and 0.749 (95%CI: 0.711-0.886), respectively.

Discussion

This study extracted the features of the cases that potentially encountered sepsis from the MIMIC-III database. Then, four ML models, SVM, RF, LR, and XGBoost were developed to make a comparison with the GRU model regarding the capability of predicting sepsis. The present study aims to demonstrate the latent possibility of utilizing GRU algorithms to conduct predictions for the HS of patients. According to the laboratory results, the GRU presents a higher AUC compared to traditional ML methods in all computed scenarios. This result implicates that the application of GRU can effectively improve the prediction performance of sepsis in ICU and it can further decrease the mortality of sepsis.

So far, sepsis is a life-threatening disease with high mortality and draws lots of attention from physicians and scientists[26]. The incidence of severe sepsis has been increasing around the world. However, the existing prediction methods of sepsis all have certain limitations. The early prediction of sepsis is a long last challenge task due to its multifactorial characteristic[27-29]. Other than that, the definition of sepsis continues to evolve because the previous definition of sepsis presents an inaccurate performance and causes a lot of wrongly identified cases. The first definition of sepsis was released in 1991 and it divided sepsis into 3 severity levels: sepsis, severe sepsis, and septic shock. Then, in 2001, this definition was updated to boost the bedside treatment of sepsis. At last, in 2016, the previous definition of sepsis has been renewed and it clarified the information of sepsis which led to accounts of prediction methods of sepsis and facilitate the prognostic and treatment of sepsis tremendously. Therefore, we selected the latest definition of sepsis, sepsis-3, as our gold standard to prove the applicability of the method.

ML method is a prevalent computational method to process large data and complex relationships between the features[30]. ML can construct a model from labeled data,

learn from the data through algorithm iteration, train the model with the data, and then apply the model to make a prediction for a certain project. In the 21st century, ML methods have been widely applied in medical prediction[24, 31]. In 2019, ML methods were used by Hidehisa Nishi [32] to predict results of patients with anterior circulation LVO who undergo mechanical thrombectomy and indicated it more accurate than previously presented pretreatment scoring systems. In 2017, a comprehensive decision support system based on ANN and Fuzzy_AHP were presented by Oluwarotimi Williams Samuel[33] to conduct heart failure risk prognostic, which achieved a high prediction accuracy using a total of 297 cases obtained from latent heart failure patients datasets. Philipp Kickingereider[34] presented an ANN based tool to make quantitative tumor response evaluation of MRI in neuro-oncology and proved it can serve as a blueprint for the application of ML methods in radiology. There also exist amounts of researches on making a prediction based on ML methods such as SVM, Random Forest, and XGBoost. However, while ML can analyze complex and divergent information, the efficiency of implementing an algorithm is reduced with the number of features involved into account. Besides, traditional ML methods cannot account for the temporal developments and real-time data assess while using recurrent neural network methods can better exploit such information. To the best of our knowledge, the prediction model with GRU deep learning methods for sepsis has never been implemented and compared with other ML methods before. For the prediction of cases with sepsis with the MIMIC-III dataset, this study indicated that the GRU model was able to perform more precisely than any other ML methods, has an AUC value of 0.801 for the testing set. Therefore, it can be concluded that GRU deep learning method has a better predictive ability for sepsis than other ML methods.

In this study, we developed a GRU deep learning algorithm as our model to predict

sepsis of the patients. The GRU cell is similar to another RNN based algorithm[35], LSTM, with some critical differences. Firstly, GRU has no hidden state while the cell state adopts the function of the hidden state from the LSTM cell design[36]. Secondly, there exists a specialized gate to determine what the cell states forget and what part of the cell state is written to, which merges the function of different gates in the LSTM structure[22, 37, 38]. The erased part of the cell state can be written to. Finally, the output of the structure is the entire cell state, which is an obvious distinction from the LSTM cell that LSTM automatically selects which part to read from the cell state to generate an output. The aforementioned differences collectively offer a new model with fewer variables than LSTM, the fewer parameters we use to conduct prediction, the sooner clinicians can provide correspondent treatment to the patients.

The GRU deep learning method model in this article is designed to conduct an accurate sepsis prognosis using clinical data readily accessible in the MIMIC-III dataset. The course of sepsis is complicated that is influenced by several kinds of features of the patients, such as vital signs, laboratory results, and demographics while the model brings to a more accurate cases assessment. Other than precision, these models have some other advantages:

1. Timely- all of the features are easy to collect in intensive care units[39, 40].
2. High generalization- After each new data is collected, the GRU deep learning model can quickly optimize its algorithm and generalize itself to the new data, which is a distinctive advantage of deep learning.
3. High practicability- GRU deep learning methods can be used together with the clinical decisions made by the clinicians, the combination of them can strongly improve the accuracy of the prediction.

Limitations

There exist some limitations in this study. Firstly, the data used to train and verify the model were all from the MIMIC-III data set, which could not guarantee that the model could be generalized to the real data of the patients around the world, and the generalization ability of the model needed to collect a lot of new data to improve. Secondly, the research we set does not take into account the impact of timing factors on the predicted results[41], in future work we will set different control groups on the different time intervals before sepsis onset. Finally, in the era of weak artificial intelligence, many deep learning algorithms are still black boxes algorithms, such as RNN. We can only control the input and output, and we can't grasp the relationship between the internal variables. It's clear that an interpretable model can better cooperate with physicians to make prognostic and treatment to the patients. In future research, we should pay attention to understanding the process of model building and make it interpretable.

Conclusion

The prediction of sepsis is a long last challenge and we proposed a novel prediction model framework for sepsis using GRU deep learning method. By concluding the result, the present study shows that the GRU deep learning method is proved to serve as a resultful predictor which utilize readily available cases for its prediction. In the future prognostic of sepsis, deep learning methods can assist clinicians to maximize the patient's opportunity of survival.

Figures

Figure1: Definition of Systemic Inflammatory Response Syndrome

Figure2: Flow chart of the inclusion in this study

Figure3: Fully gated recurrent unit

Figure4: Confusion matrix of the models in this study

- a) GRU
- b) SVM
- c) RF
- d) LR
- e) XGBoost

Figure5: AUROC performance of the models in this study

Abbreviations:

GRU: gated recurrent unit

ICU: intensive care units

MIMIC: Multiparameter intelligent monitoring in intensive care

RNN: recurrent neural network

LR: logistic regression

SVM: support vector machine

RF: random forest

XGBoost: extreme gradient boost

AUROC: area under the receiver operating

SIRS: systemic inflammatory response syndrome

SOFA: sequential organ failure assessment

ML: machine learning

ICD-9: International Classification of Diseases, ninth version

Declarations

Ethics approval and consent to participate

Research use of MIMIC III has been approved by the Institutional Review Board of the BIDMC and Massachusetts Institute of Technology.

Consent for publication

Not applicable.

Competing interests

All authors have declared that they do not have any potential conflicts of interest.

Funding

This study was funded by National Key R&D Program of China (No.2019YFCI1511200).

Authors' contributions

Xuze Zhao performed the experiment and Xuze Zhao and Bo Qu wrote the paper together.

Acknowledgements

The authors would like to thank the reviewers of this manuscript for their valuable comments, which have helped the authors improve the paper significantly.

Data Availability

MIMIC-III database is available on <http://www.healthit.cn/2018/12/26/mimic/?wpzmaction=add&postid=1187>, other data are not publicly available but are available from the corresponding author on reasonable request.

Authors' information

Xuze Zhao,¹ and Bo Qu²

¹Tianjin University; Tianjin Key Laboratory for Disaster Medicine Technology, Tianjin China.

²Tianjin University; Tianjin Key Laboratory for Disaster Medicine Technology, Tianjin China.

Correspondence should be addressed to Bo Qu; doctor_qb@tju.edu.cn

Reference

1. Angus DC, van der Poll T: **Severe Sepsis and Septic Shock**. *N Engl J Med* 2013, **369**(9):840-851.
2. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K: **The practical implementation of artificial intelligence technologies in medicine**. *Nature Medicine* 2019, **25**(1):30-36.
3. Bone RC, Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, Schein RM, Sibbald WJ: **Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine**. *Chest* 1992, **101**(6):1644-1655.
4. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R *et al*: **Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock: 2012**. *Critical Care Medicine* 2013, **41**(2):580-637.
5. Alhazzani W, Moller MH, Arabi YM, Loeb M, Gong MN, Fan E, Oczkowski S, Levy MM, Derde L, Dzierba A *et al*: **Surviving Sepsis Campaign: guidelines on the management of critically ill adults with Coronavirus Disease 2019 (COVID-19)**. *Intensive Care Medicine* 2020, **46**(5):854-887.
6. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche J-D, Coopersmith CM *et al*: **The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)**. *Jama-Journal of the American Medical Association* 2016, **315**(8):801-810.
7. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, Reinhart CK, Suter PM, Thijs LG: **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 medicine* 1996, **22**(7):707-710.
8. Taylor RA, Pare JR, Venkatesh AK, Mowafi H, Melnick ER, Fleischman W, Hall MK: **Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach**. *Academic Emergency Medicine* 2016, **23**(3):269-278.
9. Wardi G, Carlile M, Holder A, Shashikumar S, Hayden SR, Nemati S: **Predicting Progression to Septic Shock in the Emergency Department Using an Externally Generalizable Machine-Learning Algorithm**. *Annals of emergency medicine* 2021.
10. Calvert JS, Price DA, Chettipally UK, Barton CW, Feldman MD, Hoffman JL, Jay M, Das R: **A computational approach to early sepsis detection**. *Computers in Biology and Medicine* 2016, **74**:69-73.
11. Kok C, Jahmunah V, Oh SL, Zhou X, Gururajan R, Tao X, Cheong KH, Gururajan R, Molinari F, Acharya UR: **Automated prediction of sepsis using temporal convolutional network**. *Computers in Biology and Medicine* 2020, **127**.
12. Al-Mualemi BY, Lu L: **A Deep Learning-Based Sepsis Estimation Scheme**. *Ieee Access* 2021, **9**:5442-5452.
13. Johnson AEW, Pollard TJ, Shen L, Lehman L-wH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG: **MIMIC-III, a freely accessible critical care database**. *Scientific Data* 2016, **3**.
14. Deyo RA, Cherkin DC, Ciol MA: **Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases**. *Journal of clinical epidemiology* 1992, **45**(6):613-619.
15. Guo H, Li Y, Shang J, Gu M, Huang Y, Bing G: **Learning from class-imbalanced data: Review of methods and applications**. *Expert Systems with Applications* 2017, **73**:220-239.
16. Nogueira K, Penatti OAB, dos Santos JA: **Towards better exploiting convolutional neural networks for remote sensing scene classification**. *Pattern Recognition* 2017, **61**:539-556.
17. Burges CJC: **A tutorial on Support Vector Machines for pattern recognition**. *Data Mining and Knowledge Discovery* 1998, **2**(2):121-167.
18. Chang C-C, Lin C-J: **LIBSVM: A Library for Support Vector Machines**. *Acm Transactions on Intelligent Systems and Technology* 2011, **2**(3).
19. Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J: **A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial**

- prediction of landslide susceptibility.** *Catena* 2017, **151**:147-160.
20. Breiman L: **Random forests.** *Machine Learning* 2001, **45**(1):5-32.
 21. Guidotti R, Monreale A, Ruggieri S, Turin F, Giannotti F, Pedreschi D: **A Survey of Methods for Explaining Black Box Models.** *Acm Computing Surveys* 2019, **51**(5).
 22. Zheng H, Yuan J, Chen L: **Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation.** *Energies* 2017, **10**(8).
 23. Hochreiter S, Schmidhuber J: **Long short-term memory.** *Neural computation* 1997, **9**(8):1735-1780.
 24. Choi E, Schuetz A, Stewart WF, Sun J: **Using recurrent neural network models for early detection of heart failure onset.** *Journal of the American Medical Informatics Association* 2017, **24**(2):361-370.
 25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V *et al*: **Scikit-learn: Machine Learning in Python.** *Journal of Machine Learning Research* 2011, **12**:2825-2830.
 26. Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R *et al*: **Global, regional, and national causes of child mortality in 2008: a systematic analysis.** *Lancet* 2010, **375**(9730):1969-1987.
 27. Ely EW, Shintani A, Truman B, Speroff T, Gordon SM, Harrell FE, Inouye SK, Bernard GR, Dittus RS: **Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit.** *Jama-Journal of the American Medical Association* 2004, **291**(14):1753-1762.
 28. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, Rubenfeld G, Kahn JM, Shankar-Hari M, Singer M *et al*: **Assessment of Clinical Criteria for Sepsis For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).** *Jama-Journal of the American Medical Association* 2016, **315**(8):762-774.
 29. Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL: **Serial evaluation of the SOFA score to predict outcome in critically ill patients.** *Jama-Journal of the American Medical Association* 2001, **286**(14):1754-1758.
 30. He K, Zhang X, Ren S, Sun J, Ieee: **Deep Residual Learning for Image Recognition.** In: *2016 Ieee Conference on Computer Vision and Pattern Recognition.* 2016: 770-778.
 31. Muppirala UK, Honavar VG, Dobbs D: **Predicting RNA-Protein Interactions Using Only Sequence Information.** *Bmc Bioinformatics* 2011, **12**.
 32. Nishi H, Oishi N, Ishii A, Ono I, Ogura T, Sunohara T, Chihara H, Fukumitsu R, Okawa M, Yamana N *et al*: **Predicting Clinical Outcomes of Large Vessel Occlusion Before Mechanical Thrombectomy Using Machine Learning.** *Stroke* 2019, **50**(9):2379-2388.
 33. Samuel OW, Asogbon GM, Sangaiah AK, Fang P, Li G: **An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction.** *Expert Systems with Applications* 2017, **68**:163-172.
 34. Kickingreder P, Isensee F, Tursunova I, Petersen J, Neuberger U, Bonekamp D, Brugnara G, Schell M, Kessler T, Foltyn M *et al*: **Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study.** *Lancet Oncology* 2019, **20**(5):728-740.
 35. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature* 2015, **521**(7553):436-444.
 36. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J: **LSTM: A Search Space Odyssey.** *Ieee Transactions on Neural Networks and Learning Systems* 2017, **28**(10):2222-2232.
 37. Ordonez FJ, Roggen D: **Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition.** *Sensors* 2016, **16**(1).
 38. Gers FA, Schmidhuber J, Cummins F: **Learning to forget: Continual prediction with LSTM.** *Neural Computation* 2000, **12**(10):2451-2471.
 39. Nguyen HB, Rivers EP, Knoblich BP, Jacobsen G, Muzzin A, Ressler JA, Tomlanovich MC: **Early lactate clearance is associated with improved outcome in severe sepsis and septic shock.** *Critical Care Medicine* 2004, **32**(8):1637-1642.
 40. Stiell IG, Wells GA, Vandemheen KL, Clement CM, Lesiuk H, De Maio VJ, Laupacis A, Schull M, McKnight RD, Verbeek R *et al*: **The Canadian C-spine rule for radiography in alert and stable trauma patients.** *Jama-Journal of the American Medical Association* 2001, **286**(15):1841-1848.
 41. Scherpf M, Graesser F, Malberg H, Zaunseder S: **Predicting sepsis with a recurrent neural network using the MIMIC III database.** *Computers in Biology and Medicine* 2019, **113**.

Figures

Figure 1
Definition of Systemic Inflammatory Response Syndrome

Figure 2
Flow chart of the inclusion in this study

Figure 3

Fully gated recurrent unit

Figure 4

Confusion matrix of the models in this study a) GRU b) SVM c) RF d) LR e) XGBoost

Figure 5

AUROC performance of the models in this study