



# A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients

Dong Wang<sup>1,2,3†</sup>, Jinbo Li<sup>1,4†</sup>, Yali Sun<sup>1,2,3†</sup>, Xianfei Ding<sup>1,2,3†</sup>, Xiaojuan Zhang<sup>1,2,3</sup>, Shaohua Liu<sup>1,2,3</sup>, Bing Han<sup>1,2,3</sup>, Haixu Wang<sup>1,2,3</sup>, Xiaoguang Duan<sup>1,2,3</sup> and Tongwen Sun<sup>1,2,3\*</sup>

<sup>1</sup> General Intensive Care Unit, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, <sup>2</sup> Key Laboratory for Critical Care Medicine of Henan Province, Zhengzhou, China, <sup>3</sup> Key Laboratory for Sepsis of Zhengzhou, Zhengzhou, China, <sup>4</sup> Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada

## OPEN ACCESS

### Edited by:

Thippa Reddy Gadekallu,  
VIT University, India

### Reviewed by:

Abdul Rehman Javed,  
Air University, Pakistan  
Rajesh Kaluri,  
VIT University, India

### \*Correspondence:

Tongwen Sun  
fccsuntw@zzu.edu.cn

†These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Digital Public Health,  
a section of the journal  
Frontiers in Public Health

Received: 06 August 2021

Accepted: 20 September 2021

Published: 15 October 2021

### Citation:

Wang D, Li J, Sun Y, Ding X, Zhang X,  
Liu S, Han B, Wang H, Duan X and  
Sun T (2021) A Machine Learning  
Model for Accurate Prediction of  
Sepsis in ICU Patients.  
*Front. Public Health* 9:754348.  
doi: 10.3389/fpubh.2021.754348

**Background:** Although numerous studies are conducted every year on how to reduce the fatality rate associated with sepsis, it is still a major challenge faced by patients, clinicians, and medical systems worldwide. Early identification and prediction of patients at risk of sepsis and adverse outcomes associated with sepsis are critical. We aimed to develop an artificial intelligence algorithm that can predict sepsis early.

**Methods:** This was a secondary analysis of an observational cohort study from the Intensive Care Unit of the First Affiliated Hospital of Zhengzhou University. A total of 4,449 infected patients were randomly assigned to the development and validation data set at a ratio of 4:1. After extracting electronic medical record data, a set of 55 features (variables) was calculated and passed to the random forest algorithm to predict the onset of sepsis.

**Results:** The pre-procedure clinical variables were used to build a prediction model from the training data set using the random forest machine learning method; a 5-fold cross-validation was used to evaluate the prediction accuracy of the model. Finally, we tested the model using the validation data set. The area obtained by the model under the receiver operating characteristic (ROC) curve (AUC) was 0.91, the sensitivity was 87%, and the specificity was 89%.

**Conclusions:** This newly established machine learning-based model has shown good predictive ability in Chinese sepsis patients. External validation studies are necessary to confirm the universality of our method in the population and treatment practice.

**Keywords:** sepsis, machine learning, prognostication, infection, ICU patients

## INTRODUCTION

Although numerous studies and papers on sepsis are published every year, it remains a major challenge for patients and clinicians worldwide. Between 2002 and 2012, the proportion of sepsis patients admitted to hospitals in the European ICU remained unchanged; however, the severity of the disease increased significantly (1). The standardized sepsis-related mortality rate in China in 2015 was 67 deaths per 100,000, which was equivalent to more than 1 million deaths due to sepsis (2). Despite these alarming numbers, the public seems to lack an awareness about sepsis. An adult survey found that <30% of people are aware of the severity of sepsis, which was much lower than the proportion for cardiovascular diseases, cancer, and asthma (3).

To date, the diagnosis of sepsis has largely relied on determining the presence of infection and organ dysfunction (4). In addition, screening laboratory tests are often required to confirm the diagnosis. However, laboratory testing takes time, so treatment is further delayed (5).

The early detection and prediction of patients who may develop sepsis is essential to improve the adverse consequences of sepsis. Although there are many studies on the early predictions of sepsis, such as calcitonin, C-reactive protein, white blood cells, platelets, and lactic acid (6, 7). However, disappointingly, most studies are limited in clinical prediction (8). Since sepsis is a complex clinical syndrome, it contains a wide range of multifaceted clinical and biological features; therefore, a single clinical index may not be a good reflection of the disease state (9). There is still a lack of effective biomarker combinations that can distinguish patients with sepsis from those not affected with sepsis.

Current research mainly uses data collected by bedside monitors to determine the probability of sepsis in ICU patients. Bloch et al. constructed a sepsis prediction model based on the four vital signs of mean arterial pressure, heart rate, respiratory rate, and body temperature (10). The best area under the curve (AUC) was achieved with Support Vector Machine (SVM) with radial basis function, which was 88.38%. Guillén et al. used vital sign measurements and laboratory test results to predict the likelihood of severe sepsis in patients with sepsis during ICU hospitalization (11). The study showed that the AUROC based on vital signs data was 0.84; based on vital signs and laboratory results, the AUROC was calculated to be 0.882. Calvert et al. studied the correlation between pairs and triples of vital sign measurements and the overall trend (i.e., increase, decrease, and no change) of the measurements over time to predict sepsis in the adult ICU population disease (12). Their results show that the average AUROC measurement accuracy is 0.83, but requires a larger data set, which usually requires longer processing time. Since the above studies are based on the previous definition of the Third International Consensus Definition of Sepsis (sepsis-3), our current understanding of sepsis is of limited reference value. Nemati et al. used electronic medical record data combined with high-resolution time series of heart rate and blood pressure to dynamically predict sepsis, with an area under the receiver operating characteristic (AUROC) of 0.83–0.85 (13). Although the study is based on the third international definition of sepsis (sepsis-3), its predictive power is not significantly different from previous studies.

Machine learning has been applied to multiple healthcare fields, including diabetes, cancer, cardiology, and mental health (14–17). Most of the machine learning models and tools developed in the research environment has studied the potential of prognosis, diagnosis, or clinical componentization, thus demonstrating the prospect of developing computerized decision support tools (18, 19). In general, the use of machine learning

models can improve patient safety, improve the quality of care, and reduce medical costs (20).

The application of artificial intelligence in the medical field is gaining increasing recognition in the improvement of clinical practice and achievement of personalized treatment (21, 22). This study used machine learning methods to evaluate predictive clinical indicators and biomarkers related to sepsis and to establish a model that could effectively predict sepsis early, which is necessary to identify high-risk patients and may enhance the understanding and facilitate clinical management of sepsis.

## MATERIALS AND METHODS

### Study Population

This study was a secondary analysis of a retrospective observational study conducted from 2014 to 2016 in the intensive care unit (ICU) of the First Affiliated Hospital of Zhengzhou University. The inclusion criteria were (1) infection at the time of admission to the ICU; (2) compliance with the international consensus definition of sepsis and septic shock (Sepsis-3.0); (3) age  $\geq 18$  years. The exclusion criteria were (1) age  $< 18$  years; (2) diseases without infection status such as coronary heart disease, cardiac arrest, fracture, neoplasm, cerebral infarction, and brain injury; and (3) more than three missing data. Clinical or laboratory parameters related to infection and sepsis were collected for each patient.

### Statistical Analysis

The binary variables were described as counts and percentages and were evaluated using the Chi-square test or Fisher's exact test. If the continuous variables conformed to a normal distribution, they were compared using a *t*-test and expressed as means  $\pm$  SEM. For a non-normal distribution, the Mann-Whitney *U* test was used.  $P < 0.05$  was considered statistically significant. The ensemble model was written Python scripting language (Version 3.6.5, Python Software Foundation, Wilmington, DE, USA, <https://www.python.org>).

### Modeling and Feature Selection

The random forest algorithm, which belongs to the category of machine learning methods and captures non-line relationships between dependent and independent variables with high flexibility and sufficient accuracy, has been successfully applied to various fields such as the estimation of the genetic effects (23), clinical deterioration prediction (24), association estimation (25), clinical outcome prediction (26), and others (27–30). In this study, we used the random forest algorithm to predict the risk of sepsis in ICU patients by analyzing laboratory/clinic data as follows: (i) lipids, (ii) liver function, (iii) hemagglutination, (iv) blood cells, (v) renal function, and (vi) electrolyte. The essential idea of the random forest algorithm is to build multiple decision trees to reduce the correlation between trees using bootstrap aggregating or bagging, which can avoid the overfitting problem. The random forest algorithm was written in the Python scripting language (version 3.6.5, Python Software Foundation, Wilmington, DE, USA, <https://www.python.org>).

**Abbreviations:** ICU, Intensive care unit; LDL, Low-density lipoprotein; PT, Prothrombin time; ROC, Receiver operating characteristic; WBC, White blood cells.

Generally, models with more features will achieve higher accuracy than those with fewer features. However, in clinical practice, having more features cannot always improve the performance of the model because of irrelevant or redundant features, which may mislead the models. To recognize the key features and the optimal combination of features, we performed a random forest algorithm on different subsets of the training set. In this study, we identified 55 features, which were potential candidates for sepsis prediction. Because the number of possible feature combinations was large ( $2^{55}$ ), as shown in **Figure 2**, we used the Gini importance to rank the importance of all potential features (31, 32). Specifically, a high Gini importance value was a high priority for incorporation into the model. On the basis of the Gini importance value of each feature, we performed the random forest algorithm on the various feature subsets.

## Validation

In this study, we used a 5-fold cross-validation to assess the prediction performance of the model because it was the most commonly used method for machine learning-based medical problem exploration (33–37). Specifically, the available training set was divided into five roughly equal-sized subsets: the training set and the validation (or internal validation) set. Four of them were applied to fit the random forest model, and the remaining one was used to estimate the accuracy of the model.

We measured the performance of the model by applying several different indices, namely (i) AUC, (ii) accuracy, (iii) precision, (iv) recall, and (v) F1-Score, which were defined as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{TN + TP}{TN + FP + FN + TP} \\ \text{Precision} &= \frac{TP}{FP + TP} \\ \text{Recall} &= \frac{TP}{FN + TP} \\ \text{F1-Score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned}$$

Here, *TP*, *FP*, *TN*, and *FN* are the number of positive samples classified as positive (true positives), the number of negative samples classified as positive (false positives), the number of positive samples classified as negative (true negatives), and the number of negative samples classified as negative (false negatives). Briefly, we used five prediction performance indices, 5-fold cross-validation for internal validation, and the testing set for external validation to estimate the performance of the model.

## RESULTS

### Patient Characteristics

Our database consisted of 17,005 patients admitted to the ICU. After a series of exclusions, 4,449 adult patients were included in this study, and 3,539 patients developed sepsis. The process of cohort selection is shown in **Figure 1**. A total of 55 variables, including age, sex, red blood cell count, total cholesterol, D-dimer, and other clinical or laboratory parameters

related to infection and sepsis, were collected for each patient. The baseline characteristics of the included patients are shown in **Supplementary Table 1**. We then randomly divided the patients into the training and testing sets. **Supplementary Table 2** shows the basic information compared between the two sets.

### Variables of Importance

Generally, the error of the model decreased with an increase in variable selection. However, increasing the number of variables was not conducive to clinical practice. In order to identify the prominent features, we used the random forest method to select variables by using various feature subsets. Therefore, the relative importance of each feature based on the fact that the features built on the tree top contributed more to the prediction of sepsis in high-risk patients is shown in **Figure 2**. It can be observed in **Figure 3** that the error value remained at a similar degree when the number of features reached 20. Therefore, we utilized a combination of 20 selected features to predict sepsis in ICU patients (shown in **Supplementary Table 3**): neutrophils%, D-dimer, neutrophils, eosinophils, lymphocytes, albumin, white blood cells (WBCs), direct bilirubin, potassium, calcium, cholinesterase, magnesium, low-density lipoprotein (LDL), prothrombin time (PT), lymphocytes, lactate dehydrogenase (LDH), basophils%, total cholesterol (TBIL), urea, and platelets (PLT).

Next, we performed a random forest classification with the same parameters (to make the comparison possible and remove the effect of the parameters) with different subsets of features to calculate the changes in AUC values. The AUC loss value changed when we set the number of features to different values (**Supplementary Table 3**).

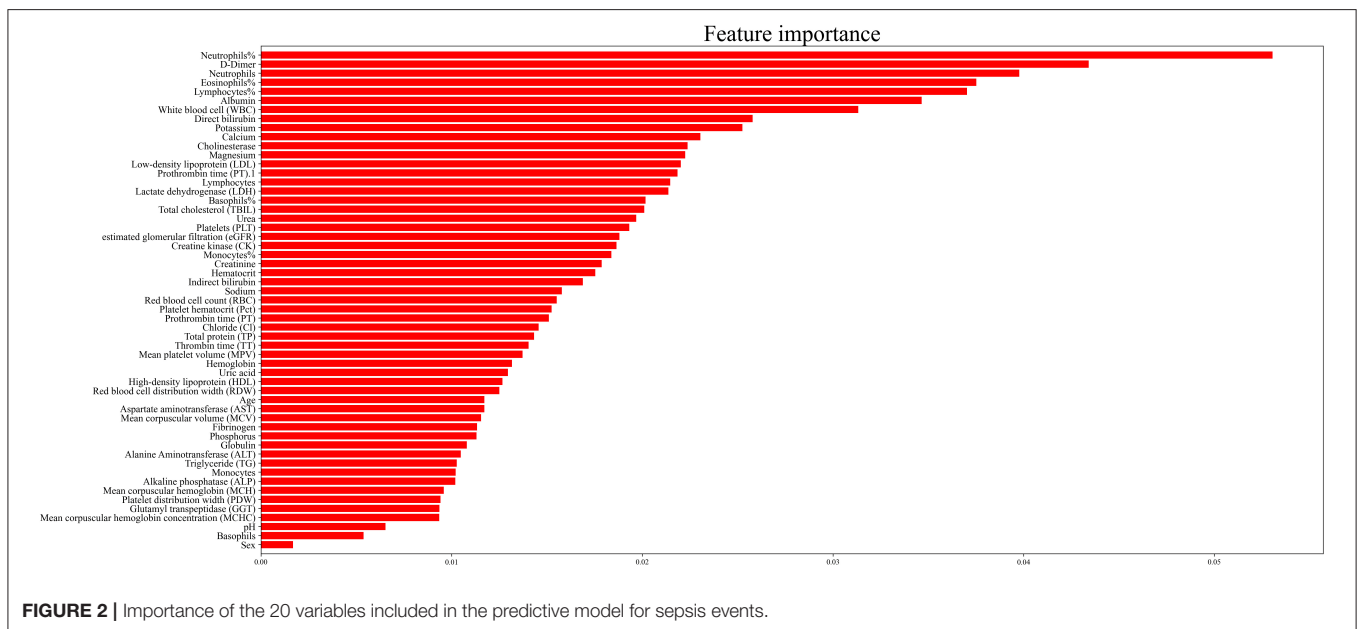
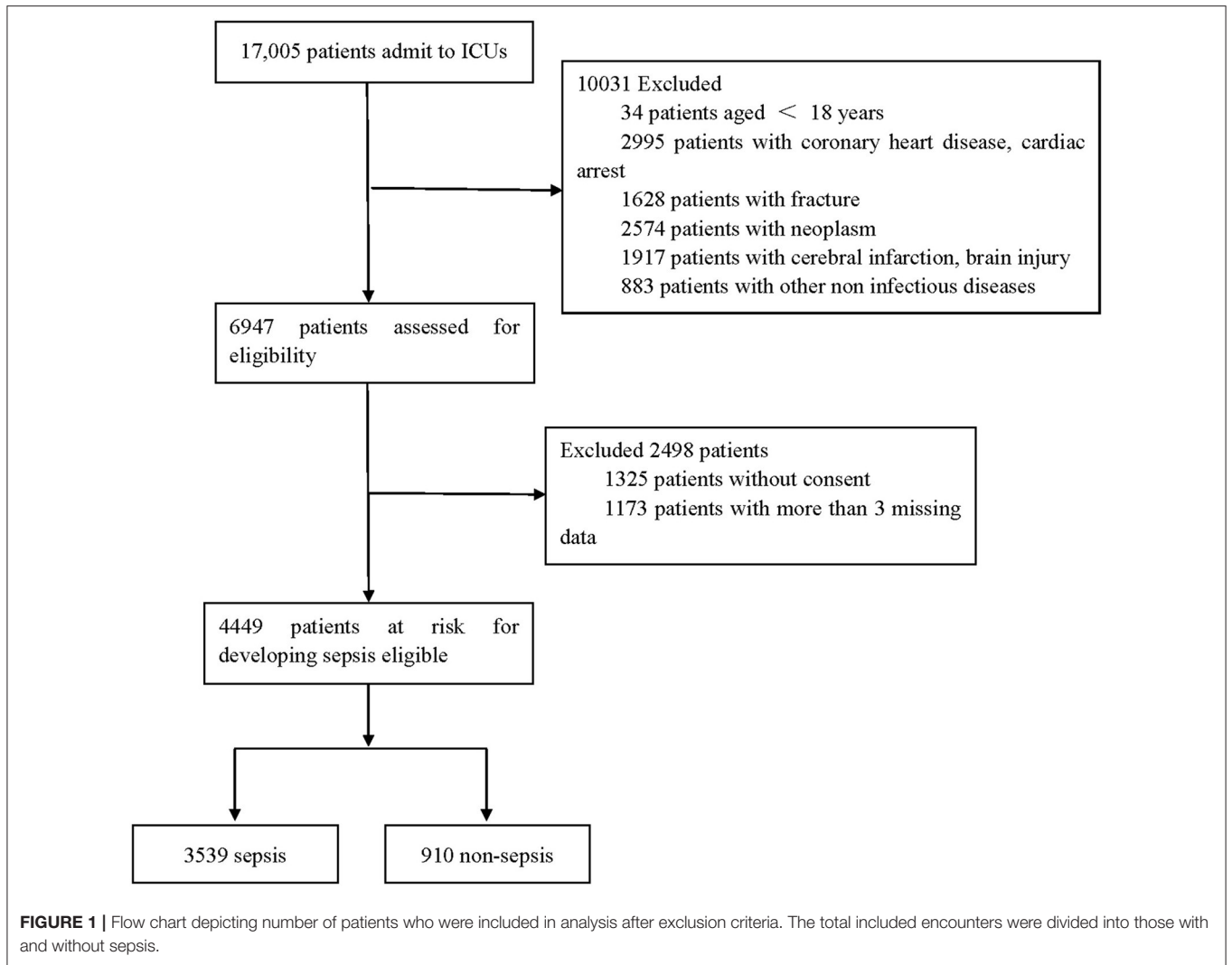
### Classification Results

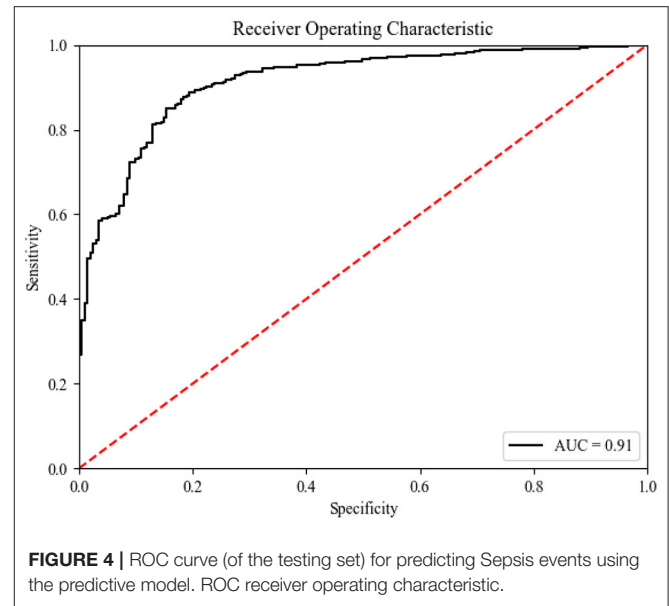
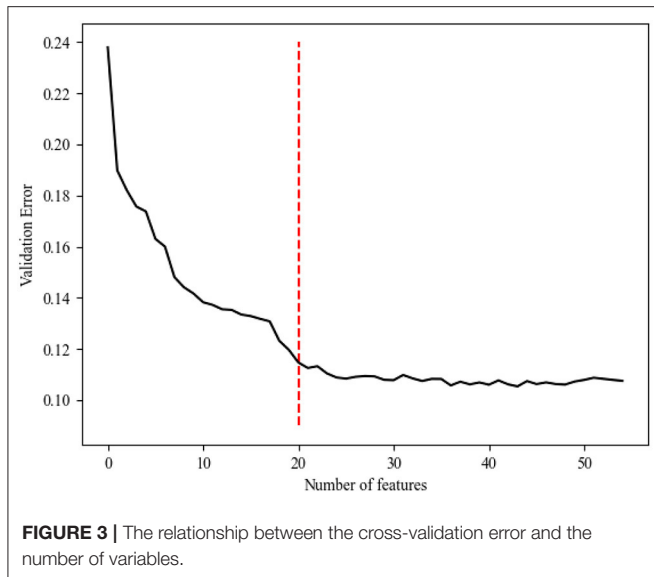
As shown in **Table 1** and **Figure 4**, on average, the random forest algorithm achieved an AUC of 0.88 ( $\pm 0.04$ ), accuracy of 0.88 ( $\pm 0.03$ ), precision of 0.90 ( $\pm 0.03$ ), recall of 0.96 ( $\pm 0.01$ ), and recall of 0.93 ( $\pm 0.02$ ) in the internal validation. For the external validation, the model gave an AUC of 0.91, accuracy of 0.87, precision of 0.89, recall of 0.95, and recall of 0.92.

## DISCUSSION

Early identification and treatment of sepsis is a highly complex and multifaceted challenge (38). It requires highly skilled and well-trained human experts (39). However, with the continuous emergence of AI applications in the medical field, some of these decisions will soon be replaced by machines called “intelligence” to improve clinical practice and patient outcomes (40). Most of what we call “artificial intelligence” is machine learning, which means learning from data and using this knowledge to acquire new knowledge or skills.

This study used a supervised learning method (a machine learning method) to build a predictive model, which included 20 predictors of sepsis events predicted by the random forest method. The AUC of this newly developed model was 0.91, demonstrating good discriminative power. These prediction





**TABLE 1 |** Internal and external validation results of the prediction model.

	<b>AUC</b>	<b>Accuracy</b>	<b>Precision</b>
Internal validation/validation set (95% CI)	0.88 (±0.04)	0.88 (±0.03)	0.90 (±0.03)
External validation/testing set	0.91	0.87	0.89
	<b>Recall</b>	<b>F1-Score</b>	
Internal validation/validation set (95% CI)	0.96 (±0.01)	0.93 (±0.02)	
External validation/testing set	0.95	0.92	

CI, confidence intervals; AUC, indicates area under the receiver-operating curves.

results suggest that the ensemble model with 20 key features is feasible and practical.

To our knowledge, most previous studies have developed models to predict the prognosis of sepsis. However, only few researchers have paid attention to the differences in the incidence of sepsis after infection, although it is important for clinical preventive intervention. Thomas et al. developed machine learning models for the early identification of sepsis risk (41); however, they did not obtain precise biomarkers that could be applied to clinicians. All calculations are trivial for a computer, which may limit generalization of the results to other hospitals and hospital systems. Other artificial intelligence systems such as random forest models may be a valuable tool to predict sepsis (8).

The variables in our model were mainly blood cells, lipids, liver function, hemagglutination, renal function, electrolyte, enzyme, and others. Interestingly, blood-related variables accounted for a large part of our model; the first five variables in **Figure 2** are related to the blood system. Neutrophils were an ideal choice for eliminating pathogenic bacteria because they store a large number of proteolytic enzymes that can rapidly produce reactive oxygen species to degrade internal

pathogens. Hence, patients with sepsis often have neutrophil infiltration, and the degree of infiltration is related to tissue damage (42, 43). Other blood cells, including eosinophils, basophils, lymphocytes, and WBCs, are also associated with the body’s defense against infection. For example, some studies have speculated that individuals with basophilic granulocytopenia have a weak resistance to infection and thus are more likely to develop sepsis (44). In addition, studies have shown that eosinophilia was a moderate marker for distinguishing SIRS from infection in critically ill patients newly admitted to the hospital, which suggested that eosinophilia may be a useful clinical tool for the prediction of sepsis (45). In addition, lymphocyte apoptosis has been recognized as an important step in the pathogenesis of experimental sepsis by inducing a state of “immune paralysis” that renders the host vulnerable to invading pathogens (46).

In the past decade, there has been a growing awareness about the role of the coagulation and fibrinolysis systems in the development of inflammation. Patients with sepsis may have common host reactions, such as coagulation, inflammation, and endothelial injury. Abnormal inflammatory and coagulation biomarkers were found to be associated with disease severity and mortality in patients with severe sepsis (47). Platelets are the main effector cells involved in blood coagulation and can promote the development of excessive inflammation, DIC, and microthrombosis (48). PT can reflect the coagulation function of the body, and D-dimer levels increase under hypercoagulable state (49). Therefore, changes in these substances may predict the occurrence of sepsis.

Sepsis is often associated with multiple organ dysfunction such as that involving the heart, liver, and kidney (50). Therefore, some indicators reflecting organ function may be used to predict the occurrence of sepsis. Albumin which is the most important protein in human plasma, maintains nutrition and osmotic

pressure. When liver synthesis is dysfunctional, its level usually decreases. Lactate dehydrogenase and urea are associated with cardiac and renal function, respectively. Patel et al. revealed an association between serum bilirubin levels and mortality during sepsis, suggesting that serum bilirubin may be a potential predictor of sepsis occurrence and death (51).

Previous studies have shown that lipids are also involved in the occurrence and development of sepsis. Yamano et al. found that low total cholesterol and high total bilirubin levels are associated with prognosis in patients with prolonged sepsis (52). Hofer et al. found that pharmacologic inhibition of cholinesterase improves survival in experimental sepsis, probably by activating the cholinergic anti-inflammatory pathway (53). The results of Feng's study suggest that a decrease in LDL-C levels is significantly associated with an increased risk of sepsis in infected patients, although the association was due to the presence of complications (54).

Although the association between electrolytes other than calcium and sepsis appears to be poorly studied, this study found that the decrease of potassium and magnesium is closely related to the occurrence of sepsis. We know that the critical illness itself is associated with a decrease in serum total calcium and free calcium levels, which is related to the severity of underlying diseases as measured by the APACHE II score. In addition, studies have shown that total and ionized hypocalcemia is more significantly associated with increased severity of infection, which suggested the role of calcium in predicting the risk of sepsis in patients with infection (55). Regarding magnesium and potassium, a study pointed out that ATP-MgCl<sub>2</sub> may be beneficial in sepsis (56). An increasing amount of evidence has suggested that potassium channels are involved in cardiovascular dysfunction in sepsis after systemic inflammation, cardiovascular dysfunction, and organ damage, and that potassium channels may affect the emergence of sepsis after infection (57). In conclusion, we believe that because sepsis is not a simple disease that can be predicted by a single marker, the biomarkers included in our model can be combined to predict the risk of sepsis in infected patients.

Our study has several limitations. First, this was a retrospective study, which had its own shortcomings, such as information bias. Second, the prediction model may have lacked generality because the 55 variables are still too few, and many other variables were omitted due to the loss of too many values. Generally, the more the variables included, the higher the

prediction accuracy. Therefore, we hope to include more patients and variables in future prospective studies.

A model with 20 key features was successfully established to predict sepsis events in Chinese patients. This model has excellent ability to predict sepsis events in Chinese patients.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

DW, YS, and JL participated in the research design and coordination and helped to draft the manuscript. XDi and XZ contributed the acquisition of data. JL, SL, BH, HW, XD, and TS performed the data analysis. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by grants from the United Fund of National Natural Science Foundation of China (U2004110); Leading Talents Fund in Science and Technology Innovation in Henan Province (194200510017); Science and Technology people-benefit project of Zhengzhou (2019KJHM0001). The special fund for young and middle-aged medical research of China International Medical Exchange Foundation (Z-2018-35); The integrated thinking research foundation of the China foundation for International Medical Exchange (Z-2016-23-2001); The study of mechanism of Gabexate Mesilate in the treatment of sepsis and septic shock (2019-hx-45).

## ACKNOWLEDGMENTS

We would like to thank Editage ([www.editage.cn](http://www.editage.cn)) for English language editing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2021.754348/full#supplementary-material>

## REFERENCES

- Vincent JL, Lefrant JY, Kotfis K, Nanchal R, Martin-Loeches I, Wittebole X, et al. Comparison of European ICU patients in 2012 (ICON) versus 2002 (SOAP). *Intensive Care Med.* (2018) 44:337–44. doi: 10.1007/s00134-017-5043-2
- Weng L, Zeng XY, Yin P, Wang LJ, Wang CY, Jiang W, et al. Sepsis-related mortality in China: a descriptive analysis. *Intensive Care Med.* (2018) 44:1071–80. doi: 10.1007/s00134-018-5203-z
- Kerrigan SW, Martin-Loeches I. Public awareness of sepsis is still poor: we need to do more. *Intensive Care Med.* (2018) 44:1771–3. doi: 10.1007/s00134-018-5307-5
- Angus DC, van der Poll T. Severe sepsis and septic shock. *N Engl J Med.* (2013) 369:840–51. doi: 10.1056/NEJMra1208623
- Herzum I, Renz H. Inflammatory markers in SIRS, sepsis and septic shock. *Curr Med Chem.* (2008) 15:581–7. doi: 10.2174/092986708783769704
- Küster H, Weiss M, Willeitner AE, Detlefsen S, Jeremias I, Zbojan J, et al. Interleukin-1 receptor antagonist and interleukin-6 for early diagnosis of neonatal sepsis 2 days before clinical manifestation. *Lancet.* (1998) 352:1271–7. doi: 10.1016/S0140-6736(98)08148-3
- Brunkhorst FM, Wegscheider K, Forycki ZF, Brunkhorst R. Procalcitonin for early diagnosis and differentiation of SIRS, sepsis, severe sepsis, and septic shock. *Intensive Care Med.* (2000) 26(Suppl. 2):S148–52. doi: 10.1007/s001340051134

8. Møller MH, Alhazzani W, Shankar-Hari M. Focus on sepsis. *Intensive Care Med.* (2019) 45:1459–61. doi: 10.1007/s00134-019-05680-4
9. Hernandez G, Bellomo R, Bakker J. The ten pitfalls of lactate clearance in sepsis. *Intensive Care Med.* (2019) 45:82–5. doi: 10.1007/s00134-018-5213-x
10. Bloch E, Rotem T, Cohen J, Singer P, Aperstein Y. Machine learning models for analysis of vital signs dynamics: a case for sepsis onset prediction. *J Healthc Eng.* (2019) 2019:5930379. doi: 10.1155/2019/5930379
11. Guillén J, Liu J, Furr M, Wang T, Strong S, Moore C, et al. Predictive models for severe sepsis in adult ICU patients. In: *2015 Systems and Information Engineering Design Symposium*. Charlottesville, VA: IEEE (2015). p. 182–7. doi: 10.1109/SIEDS.2015.7116970
12. Calvert JS, Price DA, Chettipally UK, Barton CW, Barton CW, Feldman MD, et al. A computational approach to early sepsis detection. *Comput Biol Med.* (2016) 74:69–73. doi: 10.1016/j.combiomed.2016.05.003
13. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med.* (2018) 46:547–53. doi: 10.1097/CCM.0000000000002936
14. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Comput Sci.* (2021) 7:e390. doi: 10.7717/peerj-cs.390
15. Rajput DS, Basha SM, Xin Q, Gadekallu TR, Kaluri R, Lakshmana K, et al. Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India. *J Ambient Intell Hum Comput.* (2021) 1–12. doi: 10.1007/s12652-021-03154-4
16. Javed AR, Sarwar MU, ur Rehman S, Khan HU, Al-Otaibi YD, Alnumay WS. PP-SPA: privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitively impaired individuals. *Neural Process Lett.* (2021) 1–18. doi: 10.1007/s11063-020-10414-5
17. Javed AR, Sarwar MU, Beg MO, Asim M, Baker T, Tawfk H. A collaborative healthcare framework for shared healthcare plan with ambient intelligence. *Hum Cent Comput Inf. Sci.* (2020) 10:40. doi: 10.1186/s13673-020-00245-7
18. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I, et al. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Front Neurosci.* (2015) 9:307. doi: 10.3389/fnins.2015.00307
19. Reddy GT, Reddy MPK, Lakshmana K, Rajput DS, Kalur R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intell.* (2020) 13:185–96. doi: 10.1007/s12065-019-00327-1
20. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med.* (2019) 25:433–8. doi: 10.1038/s41591-018-0335-9
21. Dhanamjayulu C, Nizhal UN, Maddikunta PR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process.* (2021) 1–12. doi: 10.1049/ipr2.12222
22. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput.* (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
23. Stephan J, Stegle O, Beyer A. A random forest approach to capture genetic effects in the presence of population structure. *Nat Commun.* (2015) 6:7432. doi: 10.1038/ncomms8432
24. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med.* (2016) 44:368–74. doi: 10.1097/CCM.0000000000001571
25. Pannaraj PS, Li F, Cerini C, Bender JM, Yang S, Rollie A, et al. Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatr.* (2017) 171:647–54. doi: 10.1001/jamapediatrics.2017.0378
26. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care.* (2019) 23:64. doi: 10.1186/s13054-019-2351-7
27. Kaup AR, NettiSimmons J, Harris TB, Sink KM, Satterfield S, Metti AL, et al. Cognitive resilience to apolipoprotein E ε4: contributing factors in black and white older adults. *JAMA Neurol.* (2015) 72:340–8. doi: 10.1001/jamaneurol.2014.3978
28. Lanspa MJ, Gutsche AR, Wilson EL, Olsen TD, Hirshberg EL, Knox DB, et al. Application of a simplified definition of diastolic function in severe sepsis and septic shock. *Crit Care.* (2016) 20:243. doi: 10.1186/s13054-016-1421-3
29. Flechet M, Güiza F, Schetz M, Wouters P, Vanhorebeek I, Derese I, et al. AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. *Intensive Care Med.* (2017) 43:764–73. doi: 10.1007/s00134-017-4678-3
30. Roimi M, Neuberger A, Shrot A, Paul M, Geffen Y, Bar-Lavie Y. Early diagnosis of bloodstream infections in the intensive care unit using machine-learning algorithms. *Intensive Care Med.* (2020) 46:454–62. doi: 10.1007/s00134-019-05876-8
31. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics.* (2009) 10:213. doi: 10.1186/1471-2105-10-213
32. Li Y, Middaugh CR, Fang J. A novel scoring function for discriminating hyperthermophilic and mesophilic proteins with application to predicting relative thermostability of protein mutants. *BMC Bioinformatics.* (2010) 11:62. doi: 10.1186/1471-2105-11-62
33. Singh J, Hanson J, Paliwal K, Zhou Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat Commun.* (2019) 10:5407. doi: 10.1038/s41467-019-13395-9
34. Mavarakis E, Ma C, Shinkai K, Fiorentino D, Callen JP, Wollina U, et al. Diagnostic criteria of ulcerative pyoderma gangrenosum: a delphi consensus of international experts. *JAMA Dermatol.* (2018) 154:461–66. doi: 10.1001/jamadermatol.2017.5980
35. Callahan M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA.* (2002) 287:2847–50. doi: 10.1001/jama.287.21.2847
36. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med.* (2018) 378:1396–407. doi: 10.1056/NEJMoa1801445
37. Cheng J, Han Z, Mehra R, Shao W, Cheng M, Feng Q, et al. Computational analysis of pathological images enables a better diagnosis of TFE3 Xp11.2 translocation renal cell carcinoma. *Nat Commun.* (2020) 11:1778. doi: 10.1038/s41467-020-15671-5
38. Helms J, Perner A. Focus on sepsis. *Intensive Care Med.* (2020) 46:1457–9. doi: 10.1007/s00134-020-06038-x
39. Komorowski M. Clinical management of sepsis can be improved by artificial intelligence: yes. *Intensive Care Med.* (2020) 46:375–7. doi: 10.1007/s00134-019-05898-2
40. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7
41. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. *JMIR Med Inform.* (2016) 4:e28. doi: 10.2196/medinform.5909
42. Edward SW. *Biochemistry and Physiology of the Neutrophil*. Cambridge: Cambridge University Press (1994).
43. Brown KA, Brain SD, Pearson JD, Edgeworth JD, Lewis SM, Treacher DF. Neutrophils in development of multiple organ failure in sepsis. *Lancet.* (2006) 368:157–69. doi: 10.1016/S0140-6736(06)69005-3
44. Piliponsky AM, Shubin NJ, Lahiri AK, Truong P, Clauson M, Niino K, et al. Basophil-derived tumor necrosis factor can enhance survival in a sepsis model in mice. *Nat Immunol.* (2019) 20:129–40. doi: 10.1038/s41590-018-0288-7
45. Abidi K, Khoudri I, Belayachi J, Madani N, Zekraoui A, Zeggwagh AA, et al. Eosinopenia is a reliable marker of sepsis on admission to medical intensive care units. *Crit Care.* (2008) 12:R59. doi: 10.1186/cc6883
46. Hotchkiss RS, Tinsley KW, Swanson PE, Schmiege RE Jr, Hui JJ, Chang KC, et al. Sepsis-induced apoptosis causes progressive profound depletion of B and CD4+ T lymphocytes in humans. *J Immunol.* (2001) 166:6952–63. doi: 10.4049/jimmunol.166.11.6952
47. Kinasevitz GT, Yan SB, Basson B, Comp P, Russell JA, Cariou A, et al. Universal changes in biomarkers of coagulation and inflammation occur in patients with severe sepsis, regardless of causative micro-organism [ISRCTN74215569]. *Crit Care.* (2004) 8:R82–90. doi: 10.1186/cc2459

48. de Stoppelaar SF, van 't Veer C, van der Poll T. The role of platelets in sepsis. *Thromb Haemost.* (2014) 112:666–77. doi: 10.1160/TH14-02-0126
49. Rodelo JR, De la Rosa G, Valencia ML, Ospina S, Arango CM, Gómez CI, et al. D-dimer is a significant prognostic factor in patients with suspected infection and sepsis. *Am J Emerg Med.* (2012) 30:1991–9. doi: 10.1016/j.ajem.2012.04.033
50. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA.* (2016) 315:801–10. doi: 10.1001/jama.2016.0287
51. Patel JJ, Taneja A, Niccum D, Kumar G, Jacobs E, Nanchal R. The association of serum bilirubin levels on the outcomes of severe sepsis. *J Intensive Care Med.* (2015) 30:23–9. doi: 10.1177/0885066613488739
52. Yamano S, Shimizu K, Ogura H, Hirose T, Hamasaki T, Shimazu T, et al. Low total cholesterol and high total bilirubin are associated with prognosis in patients with prolonged sepsis. *J Crit Care.* (2016) 31:36–40. doi: 10.1016/j.jcrc.2015.09.033
53. Hofer S, Eisenbach C, Lukic IK, Schneider L, Bode K, Brueckmann M, et al. Pharmacologic cholinesterase inhibition improves survival in experimental sepsis. *Crit Care Med.* (2008) 36:404–8. doi: 10.1097/01.CCM.0B013E31816208B3
54. Feng Q, Wei WQ, Chaugai S, Leon BGC, Mosley JD, Leon DAC, et al. Association between low-density lipoprotein cholesterol levels and risk for sepsis among patients admitted to the hospital with infection. *JAMA Netw Open.* (2019) 2:e187223. doi: 10.1001/jamanetworkopen.2018.7223
55. Müller B, Becker KL, Kränzlin M, Schächinger H, Huber PR, Nylén ES, et al. Disordered calcium homeostasis of sepsis: association with calcitonin precursors. *Eur J Clin Invest.* (2000) 30:823–31. doi: 10.1046/j.1365-2362.2000.00714.x
56. Harkema JM, Chaudry IH. Magnesium-adenosine triphosphate in the treatment of shock, ischemia, and sepsis. *Crit Care Med.* (1992) 20:263–75. doi: 10.1097/00003246-199202000-00015
57. Sordi R, Fernandes D, Heckert BT, Assreuy J. Early potassium channel blockade improves sepsis-induced organ damage and cardiovascular dysfunction. *Br J Pharmacol.* (2011) 163:1289–301. doi: 10.1111/j.1476-5381.2011.01324.x

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Li, Sun, Ding, Zhang, Liu, Han, Wang, Duan and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.