SEPRES: Sepsis prediction via a clinical data integration system and real-world studies in the intensive care unit

Authors: Chen, Qiyu, Li, Ranran, Lin, Chihche, Lai, Chiming, Chen, Dechang, Qu, Hongping, Huang, Yaling, Lu, Wenlian, Tang, Yaoqing, Li, Lei, Lu, Wenlian, Tang, Yaoqing, Li, Lei

Date: 2021-11-22T00:00:00+00:00

Abstract

Background: Sepsis is vital in critical care medicine, and early detection and intervention are key to survival. We aimed to establish an early warning system for sepsis based on a data integration system that can be implemented in the intensive care unit (ICU). Methods: We trained the LightGBM and multilayer perceptron on the open-source database Medical Information Mart for Intensive Care for sepsis prediction. An ensemble sepsis prediction model was established based on the transfer learning and ensemble learning technique on the private dataset of Ruijin Hospital. The Shapley Additive Explanations analysis was applied to present feature importance on the prediction inference. With the development of data-integrating hub to collect and transmit data from different brands of ICU medical devices, the data integration system was established to receive, integrate, standardize, and store the real-time clinical data. In this way, the sepsis prediction model developed in the ICU of the Ruijin Hospital for the real-world study of sepsis early warning on ICU management. The trial was registered with ClinicalTrials.gov (NCT05088850). Findings: Our best early warning model achieved an area under the receiver operating characteristic curve (AUC) of 0 • 9833 in the task of detecting sepsis in 4-h preceding on the open-source database, while our ensemble model achieved an AUC of 0 • 9065 0 • 9436 in the retrospective research from 1 5-h preceding on the private database, and 0 • 8636 0 • 8992 in real-time real-world studies using the data integration system in the ICU of the Ruijin Hospital. In the continuous early warning process of patients admitted to the ICU, 22 patients who met the diagnostic criteria for sepsis during hospitalization were predicted as positive cases; 29 patients without sepsis were predicted as negative cases. Additionally, 17 patients were predicted as false-positive cases; in six patients with sepsis during ICU stay, the predicted probabilities at different time nodes were all less than the warning threshold 0·7 and predicted as false-negative cases. Interpretation: Machine learning models could allow accurate and real-time inference to detect sepsis onset within 5-h preceding at most with the help of the data integration system. We identified the features such as age, antibiotics, ventilation, and net balance to be important for the sepsis prediction inference. We argue that this system has promising potential to improve ICU management by helping medical practitioners identify at-sepsis-risk patients and prepare for timely diagnosis and intervention. Funding: Shanghai Municipal Science and Technology Major Project, the ZHANGJIANG LAB, and the Science and Technology Commission of Shanghai Municipality.

Full Text

Preamble

SEPRES: Sepsis Prediction via a Clinical Data Integration System and Real-World Studies in the Intensive Care Unit

Qiyu Chen†, Ranran Li†, ChihChe Lin, Chiming Lai, Dechang Chen, Hongping Qu, Yaling Huang, Wenlian Lu*, Yaoqing Tang*, Lei Li*

Affiliations - Department of Applied Mathematics, Fudan University, Shanghai, China (Q Chen BSc, Prof W Lu PhD) - Department of Critical Care Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China (R Li PhD, D Chen MD, PhD, H Qu MD, PhD, Y Tang MD, PhD, L Li MD, PhD) - Shanghai Electric Group Co., Ltd. Central Academe, Shanghai, China (C Lin PhD, C Lai MA, Y Huang BEc)

Correspondence to: - Dr Lei Li, Department of Critical Care Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200025, China, lileiys1023@yeah.net - Dr Yaoqing Tang, Department of Critical Care Medicine, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200025, China, tangyaoqing@126.com - Dr Wenlian Lu, Department of Applied Mathematics, Fudan University, Shanghai, 200433, China, wenlian@fudan.edu.cn

† Contributed equally as first authors * Contributed equally as senior authors

Summary

Background: Sepsis is a critical condition in intensive care medicine, and early detection with timely intervention are key to patient survival. We aimed to establish an early warning system for sepsis based on a data integration system that can be implemented in the intensive care unit (ICU).

Methods: We trained LightGBM and multilayer perceptron models on the open-source Medical Information Mart for Intensive Care (MIMIC) database for sepsis prediction. An ensemble sepsis prediction model was then established using transfer learning and ensemble learning techniques on the private dataset from Ruijin Hospital. Shapley Additive Explanations analysis was applied to present feature importance for prediction inference. We developed a data-integrating hub to collect and transmit data from different brands of ICU medical devices, establishing a data integration system to receive, integrate, standardize, and store real-time clinical data. This enabled deployment of the sepsis prediction model in the ICU of Ruijin Hospital for real-world studies of sepsis early warning in ICU management. The trial was registered with ClinicalTrials.gov (NCT05088850).

Findings: Our best early warning model achieved an area under the receiver operating characteristic curve (AUC) of 0.9833 for detecting sepsis 4 hours before onset on the open-source database, while our ensemble model achieved AUCs of 0.9065–0.9436 in retrospective research from 1–5 hours preceding on the private database, and 0.8636–0.8992 in real-time real-world studies using the data integration system in the ICU of Ruijin Hospital.

During continuous early warning monitoring of ICU patients, 22 patients who met diagnostic criteria for sepsis during hospitalization were predicted as positive cases, while 29 patients without sepsis were predicted as negative cases. Additionally, 17 patients were predicted as false-positive cases. In six patients who developed sepsis during ICU stay, the predicted probabilities at different time points were all below the warning threshold of 0.7, resulting in false-negative predictions.

Interpretation: Machine learning models can enable accurate, real-time inference to detect sepsis onset up to 5 hours in advance with the support of the data integration system. We identified features such as age, antibiotics, ventilation, and net fluid balance as important for sepsis prediction inference. We argue that this system has promising potential to improve ICU management by helping medical practitioners identify at-risk patients and prepare for timely diagnosis and intervention.

Funding: Shanghai Municipal Science and Technology Major Project, the ZHANGJIANG LAB, and the Science and Technology Commission of Shanghai Municipality.

Introduction

Sepsis, an infection-induced syndrome of physiological, pathological, and biochemical abnormalities, is a global healthcare issue associated with unacceptably high mortality and long-term morbidity among ICU patients, and imposes a substantial cost burden on healthcare resources. Early detection and timely administration of appropriate antibiotics may be the most important factors for improving prognosis in sepsis patients. However, nonspecific symptoms often lead to delayed diagnosis and intervention.

Machine learning has emerged as a promising tool for early sepsis detection through intensive management based on electronic medical records, laboratory data, and biomedical signals. In 2016, Singer et al. proposed a new definition (Sepsis-3) for sepsis. Consequently, many recent sepsis prediction studies have defined sepsis using Sequential Organ Failure Assessment (SOFA) and infection criteria rather than Systemic Inflammatory Response Syndrome (SIRS). Prospective studies have shown that implementing machine learning-based sepsis prediction algorithms in hospitals can reduce in-hospital mortality and length of stay. However, many machine learning models achieve superior performance at the cost of transparency and interpretability, which has become a barrier to clinical application. Algorithms based on gradients, attention mechanisms, and Shapley values are used to interpret machine learning models.

Most sepsis detection studies have used historical medical data such as the Medical Information Mart for Intensive Care (MIMIC). However, implementing detection models in the ICU for real-time prediction is complex. The raw data needed for model inference—such as bedside data, laboratory data, demographic data, and physician orders—typically originate from different devices. Moreover, information cannot interact directly due to differences in data transfer protocols between devices. Efforts have been made to integrate bedside medical devices, but these systems integrate a limited number of devices and data types, lacking a complete clinical perspective. Furthermore, they focus primarily on data collection and presentation without additional functionality such as real-time alerts. Meanwhile, previous sepsis prediction studies were mainly retrospective, and prospective studies used only relatively simple variables, deployment methods, and models.

In this study, we aimed to develop a data integration system for the IntelliVue Information Center, ventilators, Philips ICCA system, Laboratory Information System (LIS), and Hospital Information System (HIS); an ensemble machine learning model for sepsis prediction based on Sepsis-3 criteria; and a real-time early warning system for sepsis in the ICU, named SEpsis PREdiction System (SEPRES). We deployed SEPRES in the ICU of Ruijin Hospital and conducted real-world studies to analyze its performance in ICU patient management.

Methods

SEPRES includes a data integration system equipped with a sepsis early warning module. The data integration system can collect, store, process, and display medical data through a data integration machine, physical server, and network server. The sepsis early warning module includes a sepsis prediction model and an interpretative tool. The sepsis prediction model is an ensemble of multiple machine learning models utilizing transfer learning techniques. The interpretative tool provides information on model behavior by assigning importance to input features. Our research was approved by the Ruijin Hospital Ethics Committee (No. 2020 [140]).

Medical Device Integration Hub

We developed a medical device integration hub that can acquire and transmit data from different brands of medical devices. The hub consists of customized device connection lines, a hub unit, and an integrated data receiver. An identification module containing encoding is inserted into each medical device, enabling the hub to identify the type of online device and collect data automatically according to the communication protocol. The integrated data receiver receives and translates raw data, then uploads them to the integration server through the local area network. The medical device integration hub provides the following functions: - Device online services: Detect device connections and initiate a data reading program corresponding to the device - Storage: Store parsed data in native memory - Decoding: Parse raw data into structured data for further processing - Remote Settings: Support remote system setup and send system status - Uploading: Upload received data to the specified database

Details of data extraction can be found in Appendix I.

Framework of the Data Integration System

Figure 1 [**Figure 1: see original paper**] shows the system deployment framework. The web release system of the sepsis prediction system (SEPRES) employs a browser/server architecture.

As shown in Figure 1, the system includes a physical server with a PostgreSQL database to store sepsis warning data, and the webserver deploys the system' s user access portal. The architecture can be divided into the following components: - Device side: The medical device integration hub transmits device and HIS data to the data integration system through the local area network - Data management side: Heterogeneous data are integrated into the data integration system. Interface data, service data, and model predictions are stored and managed by the Structured Query Language (SQL) server, while components needed for the sepsis early warning module are sent to the PostgreSQL database. The Message Queuing Telemetry Transport (MQTT) server sends real-time data from the data integration system to the browser - Data server side: The web server responds to browser requests and calls the sepsis early warning module. Data preprocessing and model inference are then executed, and predictions are stored in the PostgreSQL database. The data server side includes related services (real-time calculation of SOFA score, determination of suspected infection, data statistics, data charts, and historical data query) - Application side: User requests are passed to the webserver in this layer, and processing results are displayed in the system. JavaScript programs are used for dynamic HTML page development, and the AJAX interface is used for data interaction with the webserver. Spring MVC is used to build full-featured MVC modules for web applications, combined with Node.js to provide an elegant and highly maintainable method for creating templates. Users can access the system anytime and anywhere with a browser on various devices, including PCs and mobile terminals.

System Deployment

Figure 2 [**Figure 2: see original paper**] shows the medical device integration hub installed at Ruijin Hospital. The hub is placed at the bedside, receiving data from multiple devices via different interfaces shown at the bottom of the figure, storing the last 72 hours of data in native memory, and transmitting data with a time delay of less than 10 seconds. Interfaces distributed on the two sides of the hub include two universal network interfaces, four USB interfaces for mouse, keyboard, and USB drive, two HDMI and one VGA for extended display, and eight or 16 USB and Ethernet multiplexing interfaces for medical devices. The hub can integrate data from monitors, ventilators, infusion pumps, and dialysis machines. Processed data are then transmitted to the data integration system.

Sepsis Prediction Model

Our goal was to develop a sepsis prediction model that can run in real-time in hospitals. To avoid insufficient data in a specific hospital for training, we first trained models on the open-source MIMIC database and then retrained them on private hospital databases using transfer learning techniques to improve performance. The final sepsis prediction model was obtained by integrating multiple transferred models using ensemble learning techniques.

Data Acquisition

Data sources and inclusion criteria. Our study used the MIMIC-III database (version 1.4) and the private Ruijin Hospital historical (RJ) database. MIMIC encompasses approximately 40,000 patients admitted to the ICU at Beth Israel Deaconess Medical Center in Boston from 2001 to 2012. Two tasks were established: inference on the MIMIC dataset by models trained on the MIMIC dataset, and inference on the RJ dataset by models trained on both MIMIC and RJ datasets. The first task facilitates comparison with other studies, while the second enables clinical application in Ruijin Hospital.

Patients meeting all the following criteria were included in the case group: 1. At least 14 years old 2. Sepsis onset at least 5 hours after ICU admission 3. Sepsis onset is the first instance since hospital admission

Patients meeting all the following criteria were included in the control group: 1. At least 14 years old 2. ICU stay of at least 5 hours without sepsis at that time 3. Patients without ICD-9 codes for sepsis (785.52, 995.91, and 995.92) 4. SOFA score changes of no more than 1 point during any continuous 72-hour period in the ICU stay

The third criterion was excluded from the RJ database because ICD-9 codes were not recorded.

Sepsis-label definitions. Patients were followed throughout their ICU stay until discharge or development of sepsis according to the Third International Consensus for sepsis (Sepsis-3). Specifically, if the timestamps of antibiotics (t_{abx}) and blood cultures (t_{culture}) meet the condition t_{abx} - 24h ≤ t_{culture} ≤ t_{abx} + 72h, the earlier timestamp of t_{abx} and t_{culture} is defined as the timestamp of suspected infection (t_{sus}). The SOFA score was evaluated hourly within the time window [t_{onset} - 48h, t_{onset} - 24h]. The first hour with a two or more point increase in SOFA score compared to the lowest prior score is defined as the onset of sepsis (t_{onset}).

Feature extraction. We extracted 78 and 63 patient variables from the MIMIC and RJ databases, respectively. After data cleaning, we summarized these variables as features (maximum, average, median, and minimum) at hourly intervals, and missing data were padded using the nearest previous value or a preset default value. After filtering, we obtained the MIMIC dataset with 1,057 positive and 5,834 negative episodes, and the RJ dataset with 115 positive and 239 negative episodes. We used a 5-hour time window from the episodes to predict sepsis. These datasets were divided into training, validation, and test sets. See Appendix II for further details.

Prediction Model

Machine learning model. Multiple models were tested on the MIMIC dataset, including support vector machine (SVM), multilayer perceptron (MLP), gradient boosting machine (GBM), and long short-term memory (LSTM). For GBM, we used XGBoost and LightGBM as implementations. Detailed introductions to these models can be found in Appendix V.

Training method. Some redundant features were removed to accelerate SVM and MLP training. Data were standardized (i.e., each feature' s value range was linearly scaled between 0 and 1) before training to eliminate magnitude differences between features. Hyperparameters and model structures were tuned based on performance on the validation set. See Appendix VI for further details.

Transfer learning and ensemble learning. To ensure sufficient patient cases, we first trained LightGBM and MLP models on MIMIC and later transferred them to the RJ dataset in Task 2. These models were selected based on their performance in Task 1 and served as representatives of traditional machine learning and neural network models. They required retraining from Task 1 because some variables were unavailable in the RJ database.

Due to population differences between the MIMIC and RJ databases, data were standardized separately as in Task 1. Additionally, similar features (maximum, average, median, and minimum values of variables with low recording frequency at the same hour) were reduced to single features to reduce dimensionality and facilitate transfer. Model parameters were shared as initial parameters and tuned again during training on the RJ database.

Finally, we integrated the MLP and LightGBM models by averaging their outputs to make results more robust and accurate. This ensemble result served as the final output of the sepsis prediction model.

Role of the funding source. The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Sepsis Prediction Model

We evaluated our models based on accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, and specificity on the test set.

Performance on the MIMIC-III dataset. In Task 1, XGBoost and LightGBM achieved the highest AUCs among the sepsis prediction models, followed by MLP and LSTM, while SVM performed worst. Appendix VII shows the full performance of all five models. Although the GBM structure is relatively simple, it outperformed artificial neural network models. We compared our models with other models trained on the same MIMIC open-source database using Sepsis-3 criteria and reporting prediction results within 5 hours before sepsis onset, including InSight, AISE, MGP-TCN, DTW-KNN, MLA, DSPA, and MGP-AttTCN. **Table 1** shows that our models generally outperform the others. However, it should be noted that although these models all use the MIMIC-III database, differences in training and test sets due to variations in case extraction details and sepsis criteria mean the comparison is not as standardized as most machine learning benchmark comparisons.

Table 1. Results of different models on the MIMIC-III dataset

Preceding hours	Accuracy	Sensitivity	Specificity
InSight	0.9075	0.8491	0.7292
AISE	0.8462	0.9660	0.9632
MGP-TCN	~0.85	-	-
DTW-KNN	~0.88	-	-
MLA	0.982	0.746	0.9833
DSPA	0.9564	-	-
MGP-AttTCN	0.9564	-	-
LightGBM	0.9833	0.8491	0.9660

Performance on the RJ dataset. In Task 2, after transfer learning and ensemble learning, the final performance of the sepsis prediction model is shown in **Table 2** . The overall performance was similar to that of the individual LightGBM or MLP transferred models, with the ensemble AUC slightly higher than either model alone. Detailed transfer learning results can be found in Appendix VIII.

Table 2. Results of ensemble model in Task 2

Preceding hours	Accuracy	Sensitivity	Specificity
1h	0.8250	0.6667	0.9833
2h	0.9231	0.7944	0.8417
3h	0.8467	0.8639	0.9200
4h	0.9242	0.9436	0.9065
5h	0.6000	0.7083	0.9889

Feature Interpretability

We used Shapley Additive Explanations (SHAP) analysis to explore feature importance. For LightGBM models in SEPRES, antibiotics, respiratory rate, total positive end-expiratory pressure level, fibrinogen level, temperature, net fluid balance, and age were important in most models. For MLP models, age, respiratory rate, ventilation, heart rate, antibiotics, and temperature were important in most models. Some of these features (antibiotics, respiratory rate, temperature, ventilation, and heart rate) relate to the Sepsis-3 or SIRS definitions, while literature also supports associations between other features (respiratory rate, fibrinogen, net balance, and age) and sepsis severity or mortality. Detailed results are provided in Appendix IX.

SEPRES Model Inference

The detailed steps of model inference are as follows: 1. Obtain real-time patient features using SQL query statements 2. Standardize features using the scaler obtained from the training set 3. Call the trained model to obtain prediction results 4. Call the interpretive tool to obtain feature importance based on prediction results 5. Output and store prediction results and interpretations in a standard format

System Operation

SEPRES provides predictions and explanations for every ICU patient every hour, including the risk of sepsis onset in the next 5 hours, feature influence calculated by SHAP, and SOFA predictions. It has been operating at Ruijin Hospital for several months, providing hourly early warning services for over 100 ICU patients.

The PC terminal user interface is displayed in the ICU common room. **Figure 3** [**Figure 3: see original paper**] presents an example display board for all ICU patients, including predictions of sepsis onset and SOFA changes, where high and low risks are indicated by red and blue bars, respectively.

Figure 4 [**Figure 4: see original paper**] shows detailed data for a specific patient to observe current and past status.

Real-Time Performance in the ICU of Ruijin Hospital

We extracted 67 ICU stays from February 2021 to June 2021 from the system. Each stay was labeled by SOFA score changes and physician examination for infection (based on antibiotics or blood culture), with 40 stays labeled as having at least one sepsis onset at a threshold of 0.5. Data from the control group and periods near sepsis onset in the case group were included in the analysis. Statistical results are presented in **Table 3** .

Table 3. Results of real-time data

Preceding hours	Accuracy	Sensitivity	Specificity
1h	0.8243	0.8281	0.8682
2h	0.8448	0.8626	0.8802
3h	0.8483	0.8951	0.8150
4h	0.8533	0.7843	0.8106
5h	0.8583	0.7803	0.7566

Case Studies

We discuss several cases, including true-positive, true-negative, false-positive, and false-negative examples. **Figure 5** [**Figure 5: see original paper**] illustrates model predictions for positive cases near sepsis onset or negative cases over random periods (see Appendix X for details).

Figure 5. Illustrative examples of predictions. Each subplot describes the confidence index (CI) for multiple models (Y-axis) at the target time (X-axis). (i) A patient' s condition deteriorated in the early morning of February 15, 2021, with multiple organ dysfunction, and sepsis was diagnosed at 12:00 AM. Our model prediction exceeded the warning threshold of 0.7 at 9:00 AM. (ii) Despite a high SOFA score (7.0), there was no evidence of ΔSOFA ≥ 2 within 72 hours, and predictions remained below threshold. (iii) Although the patient' s SOFA score was stable at 6.0, our model made incorrect predictions. (iv) The SOFA score increased from 6.0 to 9.0 at 6:00 PM on May 9, 2021. Combined with infection evidence, the patient was diagnosed with sepsis, but the prediction remained below the warning threshold.

Our model can detect sepsis early in most cases, although a small number of false-negative and false-positive cases occur. We propose that false-negatives may result from limited collected data, leading models to produce lower predictions. Early sepsis prediction by our model effectively guided practitioners to appropriately increase attention to these patients, leading to early sepsis diagnosis and more efficient ICU patient management.

Discussion

Machine learning methods are considered promising for early sepsis warning in the ICU. Early diagnosis and timely management can effectively improve prognosis. However, sepsis may not be diagnosed promptly in clinical practice due to physician shift changes and day-night staffing rotations. Therefore, an accurate and efficient bedside early prediction system for sepsis is important.

In this study, we established SEPRES, an ICU bedside sepsis early warning system that conducts real-time sepsis prediction through a data integration system. Unlike most machine learning sepsis prediction studies using open-source databases, SEPRES was developed and performs real-time inference and analysis in the ICU of Ruijin Hospital by integrating data from the IntelliVue Information Center, ventilators, Philips ICCA system, LIS, and HIS. Additionally, the system displays patient historical data in the user interface to help physicians intuitively obtain information about patient condition changes. Although SEPRES cannot provide definitive therapeutic guidance, the probability of sepsis occurrence allows us to pay more attention to specific patients. Furthermore, weight analysis of medical factors can provide insights for therapeutic regimen selection.

To avoid the influence of insufficient data size and non-uniform data distribution across different medical centers on machine learning model training and inference, we deployed transfer learning techniques to improve performance at the specific medical center. Notably, MIMIC-III primarily enrolled white patients (40,996 of 58,976 hospital admissions), in contrast to the Ruijin Hospital private dataset composed mainly of Chinese patients, which shows significant differences in certain sepsis prediction model features (see Appendix XI). The transfer learning process improved LightGBM model prediction AUC from 0.8613–0.8913 to 0.8964–0.9348 on Ruijin Hospital historical data.

The interpretive tool may help practitioners identify risk factors. In SHAP analysis of SEPRES models, we paid special attention to insights from net fluid balance importance. As shown in Supplementary Figure 2, positive net balance indicates higher sepsis risk. Because net balance is nursing data that are difficult to collect, 4,079 out of 6,891 episodes in the MIMIC dataset lack colloid and crystalloid balance data, while only 25 out of 329 episodes in the RJ dataset lack net balance data. Therefore, net balance has not been considered a feature in most machine learning models or has been analyzed as an important sepsis prediction factor based on MIMIC datasets. Our SHAP analysis showed that negative net balance tended to decrease predicted sepsis probability. Indeed, positive cumulative fluid balance has been reported as an independent predictor of ICU mortality. Furthermore, Lin et al. showed that patients with early positive fluid balance have increased risk of developing venous thromboembolism. The weight of net balance in our SHAP analysis further emphasizes the importance of careful fluid management in critically ill patients. We argue that including net balance in the prediction model may improve both performance and clinical management in the ICU.

Our model has certain limitations. First, we enrolled only patients who remained non-septic throughout their entire ICU stay as negative controls. This enrollment condition may be too restrictive for establishing a model to predict sepsis onset, potentially causing false-positive cases. Second, as observed in consecutive case studies, patients diagnosed with sepsis shortly after ICU transfer were difficult to predict. Short data recording periods may cause false-negative cases. Finally, our model incorporates variables such as antibiotics and mechanical ventilation, making predictions influenced by physician subjective behavior.

These limitations will be addressed in future work through diverse methods, including fine-grained labeling, inclusion of data collection from the entire ICU stay, and data augmentation.

Moreover, this workflow applies to disease warnings beyond sepsis in the ICU, such as disseminated intravascular coagulation and acute kidney injury, using the data integration system to collect necessary features and data for model construction.

Contributors

WL, YT, LL, QC, RL, and Lin C conceived and designed the study. YT, LL, QC, RL, DC, HQ, and YH acquired the data. WL, Lin C, and Lai C implemented quality control of data and algorithms. WL, QC, RL, and Lai C had full access to and verified all data in the study. QC developed, trained, and applied machine learning models. Lai C developed the data integration system. YT, LL, RL, DC, and HQ performed consecutive case studies. QC and RL prepared the first draft of the manuscript. WL, LL, and YT revised the manuscript. All authors contributed to manuscript preparation.

Declaration of Interest

We report no competing interests.

Data Sharing

The MIMIC-III database can be accessed at https://physionet.org/content/mimiciii/1.4/ after becoming a PhysioNet member (https://physionet.org/). The RJ database used in this study is not publicly available. The code used to develop the model in this manuscript is available from the corresponding author upon reasonable request.

Acknowledgments

The authors acknowledge Shanghai Electric Group Co., Ltd. Central Academe for support during development of the data integration system.

Ethics Committee Approval

Our study was approved by the Ruijin Hospital Ethics Committee [ethics committee reference number: (2020) Linlunshen No. (140)].

Appendices

Appendix I: Details of Data Extraction

Serial Port. Most medical devices communicate through network or serial port connections, including RS-232 and mini-DIN. The serial port uses binary signals, so the data rate in bits per second equals the symbol rate in baud. Commonly supported bit rates range from 2400 to 115,200 bits per second.

The HL7 Interface. Device data are transmitted to the data integration system through the HL7 interface, an electronic data interchange standard for inpatient care based on the IP protocol. Using TCP/IP connections, client systems can obtain data from the interface through both active sends and queries. HL7 v2.3.1 is generally used.

Supplementary Table 1. Types of data collected from devices and systems

Source device/system	Data type	Output medium	Format
IntelliVue Information Center	Vital signs	Network	WebServices
PB 840 Ventilator	Ventilator data	RS-232	Device-specific format
Maquet Servo-i Ventilator	Ventilator data	RS-232	Device-specific format
Maquet Servo-s Ventilator	Ventilator data	RS-232	Device-specific format
Philips ICCA	Pharmacy data, GCS, urine output	Network	WebServices
Laboratory Information System	Laboratory data	Network	WebServices
Hospital Information System	Admission, discharge, hospitalization data	Network	WebServices

Appendix II: Details of Feature Extraction

Seventy-eight patient variables from MIMIC were selected as raw data. Appendix III contains the complete list. We excluded significantly incorrect records by setting variable ranges according to specialist input. When integrating the same variables from different sources, we set priorities to extract values with highest confidence. After data cleaning, these data were summarized per hour into maximum, average, median, and minimum values, except for some constant or continuous variables, totaling 285 features. Padding was used when no value existed at the corresponding time, using either the nearest previous value or the average of all patients when no valid value was available since admission. Episodes with too few valid variables were removed to ensure data quality.

We used a five-hour time window from episodes to predict sepsis, so each sample point had 1,425 features. The final dataset contained 1,057 positive and 5,834 negative episodes, divided into training, validation, and test sets. Negative episodes were divided in a 7:1:2 ratio. For positive episodes, we selected the same number as negative episodes for validation and test sets, with remaining positive episodes included in the training set. Oversampling of positive samples or downsampling of negative samples ensured a 1:1 ratio in each set.

Similar preparation steps were applied to the RJ database, though only 63 variables were available (Appendix IV). These variables were summarized hourly into 226 features. After padding and filtering, we obtained the RJ dataset with 115 positive and 239 negative episodes, divided into a training set (76 positive, 200 negative), validation set (9 positive, 9 negative), and test set (30 positive, 30 negative).

Appendix III: Complete List of Variables Used in Task 1

We consulted literature on sepsis or SOFA score prediction and selected variables mentioned in the literature that could be extracted from MIMIC-III, totaling 78 variables: MAP, heart rate, O2sat, SBP, DBP, respiratory rate, temperature, GCS, PaO2, FiO2, SpO2, cardiac output, stroke volume, stroke volume variation, tidal volume, peak inspiratory pressure, total PEEP level, O2 flow rate, WBC, hemoglobin, hematocrit, creatinine, bilirubin, direct bilirubin, platelets, INR, PTT, AST, lactate, glucose, potassium, calcium, BUN, phosphorus, magnesium, chloride, BNP, troponin I, fibrinogen, CRP, sedimentation rate, ammonia, pH, PCO2, bicarbonate, base excess, SaO2, anion gap, albumin, bands, PT, sodium, ferritin, transferrin, creatine kinase, creatine kinase-MB, LDH, troponin T, RDW, ALP, MCHC, uric acid, monocytes, lymphocytes, MCH, AaDO2, RBC, MCV, neutrophils, weight, urine output in past 24 hours, net fluid balance, ventilation, number of antibiotics in past 12, 24, and 48 hours, SOFA score, and age.

Appendix IV: Complete List of Variables Used in Task 2

Based on Appendix III, we removed variables not recorded in the Ruijin Hospital historical database and infrequently used variables based on physician input. We also added four variables involved in SOFA score calculation.

Nineteen variables were removed: GCS, O2sat, Cardiac Output, Stroke Volume, Stroke Volume Variation, Calcium, BNP, CRP, Sedimentation Rate, Ammonia, Anion Gap, Bands, Ferritin, Transferrin, Troponin T, RDW, MCHC, MCH, MCV.

Four variables were added: norepinephrine rate, epinephrine rate, dopamine rate, and dobutamine rate.

The final collection included 63 variables.

Appendix V: Machine-Learning Models

1. Support Vector Machine (SVM). SVM aims to find a hyperplane that separates data by maximizing the distance between positive data on one side and negative data on the other. For dataset {(x_i, y_i)} where y_i   {+1, -1} is the label, kernel tricks can be introduced to improve model nonlinearity by mapping data into a feature space using nonlinear mapping with kernel functions (polynomial, linear, or mixed).

2. Multi-Layer Perceptron (MLP). MLP is a classical neural network containing an input layer, several hidden layers, and an output layer. Let In denote the input layer, H_i denote hidden layers, and Out denote the output layer. Each hidden and output layer learns a nonlinear map successively: h_i = σ(W_i · h_{i-1} + b_i), where h_i is the representation at layer i, W_i and b_i are weight and bias parameters, and σ is the activation function (ReLU for hidden layers, Softmax for output). Log loss and stochastic gradient descent algorithms are used for training, with weight decay to prevent overfitting.

3. Gradient Boosting Decision Tree (GBDT). GBDT is a widely-used algorithm with effective implementations like XGBoost and LightGBM. XGBoost uses second-order loss function information for faster, better convergence compared to standard GBDT. LightGBM further accelerates training through histogram-based algorithms, gradient-based one-side sampling, and exclusive feature bundling.

4. Long Short-Term Memory (LSTM). For time-sequential data, recurrent neural networks (RNN), specifically LSTM networks, were used. LSTM networks have memory blocks with cells and gates that control information flow, enabling selective information storage and capturing long-term dependencies better than MLP or GBDT.

Appendix VI: Training Method

In Task 1, redundant features were removed to accelerate SVM and MLP training. For SVM, all average features and features with coefficient of variation > 2 were used. For MLP, only one of maximum, average, median, or minimum was kept for each laboratory variable due to low recording frequency. Data were standardized (linearly scaled between 0 and 1) before training to eliminate magnitude differences and reduce distribution differences between datasets in the next task.

We used a linear kernel with penalty factor 1 for SVM. For MLP, 460 features were selected as input using a six-layer architecture (Supplementary Table 2). We set batch size to 256, learning rate to 0.001, dropout rate to 0.6, and weight decay coefficient to 0.001. Networks were trained with the AdaGrad algorithm. For XGBoost, we set max_{depth}=6 and colsample_{bytree}=0.2. For LightGBM, we set num_{leaves}=5, lambda_{l2}=0.1, and learning_{rate}=0.2. For LSTM, we used four hidden layers with 16 one-cell memory blocks and a fully connected output layer with sigmoid activation, learning rate 0.0001 and batch size 2000.

Supplementary Table 2. MLP network structure

Layer	Units
Input	460
Layer 1	256
Layer 2	128
Layer 3	64
Layer 4	32
Layer 5	16
Output	1

Appendix VII: Results of Sepsis Prediction Models in Task 1

Supplementary Table 3. Model results in Task 1

Model	Preceding hours	Accuracy	Sensitivity	Specificity
XGBoost	1h	0.8762	0.8726	0.8711
XGBoost	2h	0.8687	0.8762	0.8736
XGBoost	3h	0.8791	0.8711	0.7123
XGBoost	4h	0.8665	0.7107	0.7123
XGBoost	5h	0.8719	0.7292	0.7437
LightGBM	1h	0.8443	0.7972	0.8113
LightGBM	2h	0.8412	0.8149	0.8189
LightGBM	3h	0.8414	0.8200	0.8090
LightGBM	4h	0.8462	0.8192	0.8255

Model	Preceding hours	Accuracy	Sensitivity	Specificity
LightGBM	5h	0.8534	0.8491	0.8797

XGBoost and LightGBM achieved the highest AUCs, followed by MLP and LSTM, with SVM performing worst. Although GBM structure is relatively simple, it outperformed artificial neural networks, possibly due to complexity leading to poor generalization. LSTM did not outperform MLP. Across models, the five prediction tasks (1–5 hours in advance) showed no significant AUC differences.

Appendix VIII: Results of Transferred Models in Task 2

In Task 2, **Supplementary Table 4** and **Supplementary Table 5** show performance of LightGBM and MLP models trained directly on the RJ dataset (top half) and with transfer learning (bottom half). Transfer-learned models performed better in most cases, indicating they can learn generic knowledge from MIMIC to improve RJ predictions. These models were ensembled by averaging.

Supplementary Table 4. LightGBM results in Task 2

Preceding hours	Transfer learning	Accuracy	Sensitivity	Specificity
1h	No	0.8500	0.7333	0.9667
1h	Yes	0.9183	0.9311	0.9000
2h	No	0.7292	0.5889	0.8778
2h	Yes	0.9348	0.9171	1.0000
3h	No	0.7600	0.5833	0.8750
3h	Yes	0.8964	0.7056	0.9833
4h	No	0.7556	0.5733	0.9467
4h	Yes	0.8613	0.5800	0.9889
5h	No	0.8167	0.5889	0.9222
5h	Yes	0.8913	0.7333	0.9889

Supplementary Table 5. MLP results in Task 2

Preceding hours	Transfer learning	Accuracy	Sensitivity	Specificity
1h	No	0.7750	0.5667	0.9833
1h	Yes	0.7900	0.6133	0.9667
2h	No	0.7778	0.5889	0.9667
2h	Yes	0.8000	0.6222	0.9778
3h	No	0.8083	0.6333	0.9833
3h	Yes	0.8083	0.6500	0.9667
4h	No	0.9150	0.9833	0.9889
4h	Yes	0.7944	0.6000	0.9889

Preceding hours	Transfer learning	Accuracy	Sensitivity	Specificity
5h	No	0.9037	0.9667	0.9750
5h	Yes	0.8333	0.6917	0.9667

Appendix IX: Feature Importance

We used SHAP analysis to explore feature importance across models. The SHAP value represents each feature’ s impact on model output. For LightGBM models trained on MIMIC-III, beeswarm plots in **Supplementary Figure 1** show the twenty most important features combined across hours, including FiO2, fibrinogen, calcium, ventilation, age, albumin, antibiotics, O2Sat, PCO2, peak inspiratory pressure, and SOFA.

For SEPRES models, **Supplementary Figure 2** and **Supplementary Figure 3** show feature importance for LightGBM and MLP models. While both achieved high AUCs, feature importance differed. LightGBM emphasized antibiotics, respiratory rate, total PEEP level, fibrinogen, temperature, net balance, and age. MLP emphasized age, respiratory rate, ventilation, heart rate, antibiotics, and temperature. Some features relate to Sepsis-3/SIRS definitions, while literature supports associations between others (respiratory rate, fibrinogen, net balance, age) and sepsis severity/mortality.

Appendix X: Case Studies

In Ruijin Hospital operation, the threshold was increased from the default 0.50 to 0.70 to reduce false alarm rates. This adjusted specificity to ~0.88 while reducing sensitivity to ~0.72 (**Supplementary Table 6**).

Supplementary Table 6. Real-time data results with threshold at 0.70

Preceding hours	Accuracy	Sensitivity	Specificity
1h	0.7480	0.6930	0.8861
2h	0.7770	0.7351	0.8855
3h	0.7716	0.7265	0.8912
4h	0.7747	0.7340	0.8852
5h	0.7855	0.7475	0.8909

Positive case: On February 14, 2021, a 49-year-old man was transferred to the ICU after pancreaticoduodenectomy due to intestinal obstruction and abdominal infection. After drainage surgery, his condition deteriorated on February 15 morning with multiple organ dysfunction. At 12:00 AM, SOFA score was 13 (ΔSOFA ≥ 2 within 72 hours), diagnosing sepsis per Sepsis-3. Our model predicted sepsis three hours in advance, with confidence index (CI) of 0.7161 at 9:00 AM, exceeding the 0.70 threshold (**Supplementary Figure 4**). High

CI at 10:00 AM and 11:00 AM indicated deterioration. Early prediction guided practitioners to appropriately increase attention, leading to early sepsis diagnosis.

Supplementary Figure 4. Confidence index (CI) of sepsis prediction for the patient at each time point.

Negative case I: On May 20, 2021, a 45-year-old man was admitted for chronic renal failure (uremia with acute exacerbation) with metabolic acidosis and renal failure after transplantation plus severe hypertension. Despite high SOFA score (7.0) and severe condition, no ΔSOFA ≥ 2 occurred within 72 hours. CI remained below 0.70 threshold (**Supplementary Figure 5**), correctly indicating no sepsis.

Supplementary Figure 5. Confidence index (CI) of sepsis prediction for the patient at each time point.

Negative case II: On May 11, 2021, a 58-year-old man was admitted for pancytopenia with bloodstream infection, respiratory failure, chronic kidney disease, severe hypertension and diabetes. After MDT discussion, infection-induced myelosuppression was diagnosed and anti-infectious treatment given. The patient was septic on admission but improved in ICU with stable vital signs and no new ΔSOFA ≥ 2 within 72 hours. CI remained below 0.70 threshold (**Supplementary Figure 6**), correctly indicating no new sepsis occurrence.

Supplementary Figure 6. Confidence index (CI) of sepsis prediction for the patient at each time point.

Negative case III: On May 20, 2021, a 69-year-old man was admitted for cardiac insufficiency and pneumonia. After anti-infectious therapy and organ function maintenance, condition improved with stable vital signs. SOFA score remained stable at 5.0 with no ΔSOFA ≥ 2 within 72 hours, and CI remained below 0.70 threshold (**Supplementary Figure 7**), correctly indicating no sepsis.

Supplementary Figure 7. Confidence index (CI) of sepsis prediction for the patient at each time point.

False positive case I: On May 1, 2021, an 84-year-old man was admitted for intracranial space-occupying lesion (frontotemporal malignant tumor). He was transferred to ICU for hospital-acquired pneumonia (Klebsiella) and respiratory failure after surgery. Despite severe condition with temperature fluctuations, SOFA score remained stable at 6.0 with no ΔSOFA ≥ 2 within 72 hours, indicating no sepsis. However, CI for sepsis incidence within 5 hours exceeded 0.70 threshold (**Supplementary Figure 8**), creating a false positive.

Supplementary Figure 8. Confidence index (CI) of sepsis prediction for the patient at each time point.

False positive case II: On January 6, 2021, a 26-year-old man was transferred to ICU for cardiogenic shock, hyperthyroid heart disease, and pneumonia. Diagnosed with sepsis on admission, he received ventilation and anti-infectious therapy. During monitoring, despite severe condition and significant temperature fluctuations, SOFA score remained stable at 7.0 with no ΔSOFA ≥ 2 within 72 hours, and condition improved after infection control. However, CI exceeded 0.70 threshold (**Supplementary Figure 9**), creating a false positive.

Supplementary Figure 9. Confidence index (CI) of sepsis prediction for the patient at each time point.

The negative control group used during training consisted of non-septic patients, so the model may falsely identify severely ill patients as septic. This limitation will be addressed by stratifying control groups by severity and optimizing the prediction model.

False negative case I: On May 1, 2021, a 63-year-old man was admitted for sellar tumor. On May 8, during transnasal transsphenoidal resection, he experienced hemorrhagic shock and hypoxic-ischemic encephalopathy. Transferred to ICU with ventilation and anti-infectious therapy, SOFA score increased from 6.0 to 9.0 (ΔSOFA ≥ 2 within 72 hours) at 6:00 PM on May 9, diagnosing sepsis. However, CI remained below 0.70 threshold (**Supplementary Figure 10**), creating a false negative.

Supplementary Figure 10. Confidence index (CI) of sepsis prediction for the patient at each time point.

False negative case II: On March 11, 2021, a 46-year-old man was admitted for severe acute pancreatitis with abdominal infection. During monitoring, obvious temperature fluctuations occurred, and SOFA score increased significantly (ΔSOFA ≥ 2 within 72 hours) at 6:00 AM on March 13, indicating sepsis. However, CI remained below 0.70 threshold (**Supplementary Figure 11**), creating a false negative.

Supplementary Figure 11. Confidence index (CI) of sepsis prediction for the patient at each time point.

False negative case III: On March 15, 2021, a 55-year-old man was admitted for severe acute pancreatitis. After debridement and drainage on March 22, he was transferred to ICU for anti-infectious therapy. During monitoring, obvious temperature fluctuations occurred, and SOFA score increased significantly (ΔSOFA ≥ 2 within 72 hours) at 10:00 AM on March 23, indicating sepsis. However, CI remained below 0.70 threshold (**Supplementary Figure 12**), creating a false negative.

Supplementary Figure 12. Confidence index (CI) of sepsis prediction for the patient at each time point.

All three false negative cases had relatively short ICU duration, limiting collected data for prediction and potentially causing inaccurate sepsis prediction. Supporting evidence: only 17 positive cases in the RJ historical dataset had onset within 2 days, which may be insufficient for the model to learn positive patterns with limited data.

As a further experiment, we selected RJ database cases screened out due to too few valid variables. Performance on these cases (**Supplementary Table 7**) showed substantial sensitivity decrease with only slight specificity decrease, suggesting insufficient valid data leads to lower CI and false negatives. This may require inclusion of more positive cases with fewer valid variables and data augmentation.

Supplementary Table 7. Results on cases with few valid variables

Preceding hours	Accuracy	Sensitivity	Specificity
1h	0.6325	0.3150	0.9500
2h	0.6283	0.3167	0.9400
3h	0.6338	0.3300	0.9375
4h	0.6350	0.3180	0.9520
5h	0.6283	0.3050	0.9517

Appendix XI: Differences in Features Between MIMIC and RJ Datasets

Supplementary Figure 13 [Figure 13: see original paper] shows feature distributions with relatively large differences between MIMIC and RJ datasets. Each subplot displays density functions of a feature on both datasets obtained by kernel density estimation using Gaussian kernel functions.

References

1. Cecconi M, Evans L, Levy M, Rhodes A. Sepsis and septic shock. Lancet 2018; 392: 75–87.
2. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016; 315: 801–10.
3. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. Crit Care Med 2001; 29: 1303–10.
4. Marik PE, Farkas JD. The changing paradigm of sepsis: early diagnosis, early antibiotics, early pressors, and early adjuvant treatment. Crit Care Med 2018; 46: 1690–2.
5. Filbin MR, Lynch J, Gillingham TD, et al. Presenting symptoms independently predict mortality in septic shock: Importance of a previously unmeasured confounder. Crit Care Med 2018; 46: 1592–9.

6. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015; 7: 299ra122.
7. Lauritsen SM, Kalør ME, Kongsgaard EL, et al. Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med 2020; 104: 101820.
8. Desautels T, Calvert J, Hoffman J, 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: e5909.
9. 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.
10. Moor M, Horn M, Rieck B, Roqueiro D, Borgwardt K. Early recognition of sepsis with Gaussian process temporal convolutional networks and dynamic time warping. Machine Learning for Healthcare Conference; 2019: PMLR.
11. Barton C, Chettipally U, Zhou Y, et al. Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. Comput Biol Med 2019; 109: 197–205.
12. Asuroglu T, Ogul H. A deep learning approach for sepsis monitoring via severity score estimation. Comput Methods Programs Biomed 2021; 198: 105816.
13. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual 2017; 6: e000158.
14. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: A randomised clinical trial. BMJ Open Respir Res 2017; 4: e000234.
15. Zhang D, Yin C, Hunold KM, Jiang X, Caterino JM, Zhang P. An interpretable deep-learning model for early prediction of sepsis in the emergency department. Patterns (N Y) 2021; 2: 100196.
16. Rosnati M, Fortuin V. MGP-AttTCN: An interpretable machine learning model for the prediction of sepsis. PLoS One 2021; 16: e0251248.
17. Oei SP, van Sloun RJ, van der Ven M, Korsten HH, Mischi M. Towards early sepsis detection from measurements at the general ward through deep learning. Intell Based Med 2021; 5: 100042.
18. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016; 3: 1–9.
19. Sorani MD, Hemphill JC, Morabito D, Rosenthal G, Manley GT. New approaches to physiological informatics in neurocritical care. Neurocrit Care 2007; 7: 45–52.
20. Goldstein B, McNames J, McDonald BA, et al. Physiologic data acquisition system and database for the study of disease dynamics in the intensive care unit. Crit Care Med 2003; 31: 433–41.

21. Sun Y, Guo F, Kaffashi F, Jacono FJ, DeGeorgia M, Loparo KA. INSMA: An integrated system for multimodal data acquisition and analysis in the intensive care unit. J Biomed Inform 2020; 106: 103434.
22. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016: 785–94.
23. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems 2017; 30: 3146–54.
24. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems 2017: 4768–77.
25. Kenzaka T, Okayama M, Kuroki S, et al. Importance of vital signs to the early diagnosis and severity of sepsis: Association between vital signs and sequential organ failure assessment score in patients with sepsis. Intern Med 2012; 51: 871–6.
26. Matsubara T, Yamakawa K, Umemura Y, et al. Significance of plasma fibrinogen level and antithrombin activity in sepsis: A multicenter cohort study using a cubic spline model. Thromb Res 2019; 181: 17–23.
27. Brotfain E, Koyfman L, Toledano R, et al. Positive fluid balance as a major predictor of clinical outcome of patients with sepsis/septic shock after ICU discharge. Am J Emerg Med 2016; 34: 2122–6.
28. Martin GS, Mannino DM, Moss M. The effect of age on the development and outcome of adult sepsis. Crit Care Med 2006; 34: 15–21.
29. Burdick H, Pino E, Gabel-Comeau D, et al. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: A prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. BMJ Health Care Inform 2020; 27: e100109.
30. Lin TL, Dhillon NK, Conde G, et al. Early positive fluid balance is predictive for venous thromboembolism in critically ill surgical patients. Am J Surg 2021; 222: 220–6.

Research in Context

Evidence before this study. We searched PubMed from inception to September 30, 2021, using keywords "machine learning" or "deep learning" or "artificial intelligence" and "sepsis," and keywords "critical care" or "ICU" or "sepsis" and "data integration system" or "data acquisition system" or "integrated," without language restrictions. Previous studies built various prediction models based on different sepsis definitions and datasets, or built data acquisition systems integrating some data types. However, most did not combine both approaches for real-time prediction with detailed process description.

Added value of this study. We developed a complete sepsis prediction system for the ICU, including practical procedures. Machine learning models achieved high AUC scores on both databases, and we interpreted predictions. We examined our system through consecutive case studies. This workflow is also applicable to other ICU disease warnings beyond sepsis.

Implications of all available evidence. Our system can display real-time patient conditions, identify patients likely to develop sepsis, help practitioners focus attention, and support future research.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.