Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(7): 13356–13378.
DOI: 10.3934/mbe.2023595
Received: 16 March 2023
Revised: 30 May 2023
Accepted: 05 June 2023
Published: 09 June 2023

Research article

Temporal convolution attention model for sepsis clinical assistant diagnosis prediction

Yong Li and Yang Wang*

College of Computer Science and Engineering, Northwest Normal University, 967 Anning East Road, Lanzhou 730070, China

* **Correspondence:** Email: 2021222165@nwnu.edu.cn; Tel: +8619119390376.

Abstract: Sepsis is an organ failure disease caused by an infection acquired in an intensive care unit (ICU), which leads to a high mortality rate. Developing intelligent monitoring and early warning systems for sepsis is a key research area in the field of smart healthcare. Early and accurate identification of patients at high risk of sepsis can help doctors make the best clinical decisions and reduce the mortality rate of patients with sepsis. However, the scientific understanding of sepsis remains inadequate, leading to slow progress in sepsis research. With the accumulation of electronic medical records (EMRs) in hospitals, data mining technologies that can identify patient risk patterns from the vast amount of sepsis-related EMRs and the development of smart surveillance and early warning models show promise in reducing mortality. Based on the Medical Information Mart for Intensive Care III, a massive dataset of ICU EMRs published by MIT and Beth Israel Deaconess Medical Center, we propose a Temporal Convolution Attention Model for Sepsis Clinical Assistant Diagnosis Prediction (TCASP) to predict the incidence of sepsis infection in ICU patients. First, sepsis patient data is extracted from the EMRs. Then, the incidence of sepsis is predicted based on various physiological features of sepsis patients in the ICU. Finally, the TCASP model is utilized to predict the time of the first sepsis infection in ICU patients. The experiments show that the proposed model achieves an area under the receiver operating characteristic curve (AUROC) score of 86.9% (an improvement of 6.4%) and an area under the precision-recall curve (AUPRC) score of 63.9% (an improvement of 3.9%) compared to five state-of-the-art models.

Keywords: sepsis assistant diagnosis; early prediction; electronic medical records; temporal convolutional network; attention mechanism

1. Introduction

In recent years, sepsis, a severe and complex infectious disease, has imposed a tremendous burden on global healthcare systems. Accurately predicting the occurrence and progression of sepsis is crucial for timely intervention and treatment, aiming to reduce patient suffering and improve survival rates. It is one of the leading causes of death for patients in ICUs, characterized by tissue damage, organ dysfunction, and ultimately death, triggered by infection [1]. Over the past decade, approximately 30 million people worldwide (including 4.2 million newborns and children) have suffered from sepsis, and approximately 6 million patients have died. The average annual incidence of sepsis is 437 cases per $100,000$ people [2,3]. Severe sepsis exacerbates multiple organ failure [4], with a mortality rate of 9.7% [5]. Especially in patients with infectious shock symptoms, the mortality rate can increase to over 40% [6]. Studies have shown that for every hour of delay in physicians making diagnostic conclusions, the mortality rate of patients increases by 7.6% [7]. However, there is currently a lack of an effective method to address this medical problem. Developing an intelligent diagnostic early warning model to identify the risk of contracting sepsis and recommend timely therapeutic measures would significantly reduce the risk of patient death.

Predicting the incidence of sepsis is a rigorous scientific challenge in medical diagnosis. It requires empirical analysis and experience summary based on real patient EMRs to provide data sources for subsequent clinical assistant diagnosis and treatment evaluation. Despite the widespread application of artificial intelligence techniques in the field of smart healthcare [8–10], in real-world medical scenarios, EMR data is complex in structure, and diagnostic information carries a high degree of uncertainty. Therefore, current existing methods still face challenges in accurately predicting the onset of sepsis and cannot be widely applied for clinical assistant diagnosis.

The main challenge in accurately predicting the incidence of sepsis in ICU patients based on EMRs includes the following two aspects. First, the presence of numerous time-series features in EMRs complicates modeling, analysis, and clinical application. Time-series features in EMRs are crucial for early warning sepsis surveillance. However, the uncertainty of physiological data and laboratory test results at specific time intervals complicates modeling and analysis. Existing methods often utilize data from a single time point, disregarding the time-series nature of the data. Although recurrent neural network (RNN) technologies are commonly used for modeling time-series data [11], the RNN structure can suffer from issues like gradient disappearance and gradient explosion, particularly when dealing with longer time intervals, resulting in poor prediction outcomes. Therefore, it is necessary to construct a sepsis prediction model that incorporates time-series features for clinical assistant diagnosis.

Second, there were numerous missing values in the EMRs. The problem of missing data is primarily due to doctors' failure to record specific patient information or data corruption caused by other reasons. However, the issue of missing values in the data may also be caused by the variability of practice guidelines in daily medical management. For example, for patients with stable conditions, their respiratory rate, heart rate, body temperature, and other physiological signs are usually recorded every 3 to 4 hours. However, for patients in the ICU, these signs must be recorded every hour [12]. These variations in data collection intervals contribute to missing values in the EMRs, which further complicates the prediction of sepsis incidence.

Traditionally, researchers have adopted the method of deleting or averaging to solve the problem of missing data. However, these methods do not consider the uncertainty of medical data, which may

lead to severe prediction bias or invalid diagnostic conclusions.

In this study, we propose a temporal convolution attention model for sepsis clinical assistant diagnosis prediction (TCASP). The contributions of this study are summarized as follows:

1) We propose the TCASP model, which inputs multivariate time-series features into a temporal convolutional network (TCN) and combines the attention mechanism. Based on the TCASP model, we solved the modeling and clinical application problems of time-series data in a sepsis prediction model so that the physiological sign data and laboratory test results of patients' time-series nature can be fully mined and applied. The TCASP model effectively improved the prediction accuracy of sepsis and reduced the mortality of patients with sepsis.

2) We model the covariance of the EMRs and use the multitask Gaussian process (MGP) to generate a continuous function fitting for irregular sample data to fill in the missing values in the EMRs, effectively solved the uncertainty of the data in the prediction of sepsis.

3) We conducted numerous experiments on the international open source real EMRs dataset MIMIC-III. Experiments have shown that the TCASP model is superior to traditional machine learning and deep learning models in terms of prediction accuracy, effectiveness, and stability.

The remainder of this paper is organized as follows. In Section 2, we present related work. Section 3 focuses on the methodology and discusses the details of the TCASP model. Section 4 presents the results of our empirical evaluation of TCASP using the MIMIC-III dataset. Finally, we discuss concluding remarks and research directions in Section 5.

2. Related work

2.1. Algorithm for early detection of sepsis

In recent years, various methods have been proposed for detecting sepsis in ICU [13–17]. These methods have been selectively compared to the simple clinical scoring criterion known as systemic inflammatory response syndrome (SIRS) [18]. However, the SIRS clinical scoring criterion does not provide the capability for continuous assessment of sepsis risk. In addition to SIRS, sequential organ failure assessment (SOFA) score [19] and modified early warning score (MEWS) [20] are commonly used in clinical practice. These methods involve manual tabulation and analysis of patients' vital signs and laboratory results to generate risk scores and have been validated for sepsis detection in multiple studies [21–23]. However, the effectiveness of these scoring systems is limited, partly because they do not consider the temporal trends in patient data and correlations between measurements. Henry et al. [24] introduced a targeted real-time warning score system called TEWScore for predicting septic shock. Horng et al. [25] developed automated triggers using machine learning techniques in the triage process of the emergency department to support clinical decision-making for sepsis. Nemati et al. [26] proposed an interpretable machine learning model aimed at improving the predictive accuracy of sepsis and providing explanations for the prediction results. Although these machine learning methods have demonstrated superior predictive performance compared to existing simple clinical approaches, they have not adequately addressed the issue of long time-series features in clinical medical data.

2.2. Time series prediction

Time series prediction is a method that utilizes historical data of past observations to predict future time points or intervals, and it finds wide applications in various fields [27–29]. Within the context of time series and sequential data modeling, recurrent neural networks (RNNs) have been recognized as commonly used approaches due to their powerful predictive capabilities and ability to capture variable-length observations and long-term temporal dependencies. Several RNN-based architectures have been developed for sepsis prediction in longitudinal medical records, where comprehensive medical records of the same patient at different time points are considered. For instance, Kam et al. [11] leveraged the commonly used LSTM architecture in deep learning to learn sequential patterns from medical data, thereby enhancing feature extraction performance and applying it to sepsis onset prediction. Lin et al. [30] proposed a convolutional-LSTM model that combines static and dynamic information for early diagnosis and prediction of sepsis. Baral et al. [31] introduced an enhanced bidirectional LSTM network for early prediction of cardiac arrest in sepsis patients. Rafiei et al. [32] presented a novel technique called smart sepsis predictor (SSP) to forecast sepsis onset in ICU patients. Saqib et al. [33] combined traditional machine learning techniques with LSTM to enable early prediction of sepsis in electronic medical records. However, the long-term temporal dependencies inherent in the RNN structure can lead to the vanishing or exploding gradient problems, which subsequently degrade the predictive performance of RNN models.

2.3. Attention-based neural networks

Bahdanau et al. [34] first applied the attention mechanism to natural language processing to achieve simultaneous alignment and translation, solving a potential problem of the previous use of the encoder–decoder architecture in machine translation, that is, the information is compressed in a fixed-length vector that cannot correspond to long sentences. With the continuous development of attention mechanisms, various models incorporating attention have been proposed and applied in the healthcare field. By focusing limited attention on crucial information, attention mechanisms can save resources and rapidly acquire the most effective information. Consequently, it has been demonstrated that attention mechanisms significantly enhance model performance. For example, Choi et al. [35] proposed a reverse temporal attention model (RETAIN), which generates attention weights through a reverse recurrent network for detecting important clinical variables; Usama et al. [36] proposed a self-attention-based recurrent convolutional neural network (RCNN) model for assisting disease diagnosis by automatically learning advanced semantic features from clinical text using bidirectional recurrent connections in convolutions. However, the mentioned model utilized an RNN architecture that is slow to train and suffers from the vanishing gradient problem. Lan et al. [37] introduced an end-to-end computational model based on graph attention network for lncRNA-disease association prediction (GANLDA). Lin et al. [38] proposed a hierarchical attention temporal convolutional network (HA-TCN) architecture for diagnosing myotonic dystrophy from grip force time-series data, but such methods for disease prediction only focus on time points and do not consider different features. The attention-based mechanism of the TCASP proposed in this study can address these problems.

3. Methods

3.1. Problem definition

Let the set of ICU patients $H = \{h_1, h_2, \cdots, h_n\}$, $h_t = \{Z_t | temp_k, heartrate_k, spo2_k, \cdots\}$, and the feature vector Z_t represent the ICU monitoring information, such as "body temperature", "heart rate" and "pulse oximetry", of ICU patient h at time point t. The patient's stay in ICU is a continuous time interval T, $T = \{t_1, t_2, \cdots, t_j\}$. Subsequently, we let $Z_c = \bigcup_{t=1}^{j} Z_t$ denote the multivariate time series features vector Z of patient c at all time points during hospitalization. The problem of sepsis prediction can be formalized as follows: given the multivariate time-series features vector Z of T in ICU patient h during hospitalization, the decision goal is to calculate the probability p_h =TCASP(h, T, Z) of sepsis in patient h through the TCASP model.

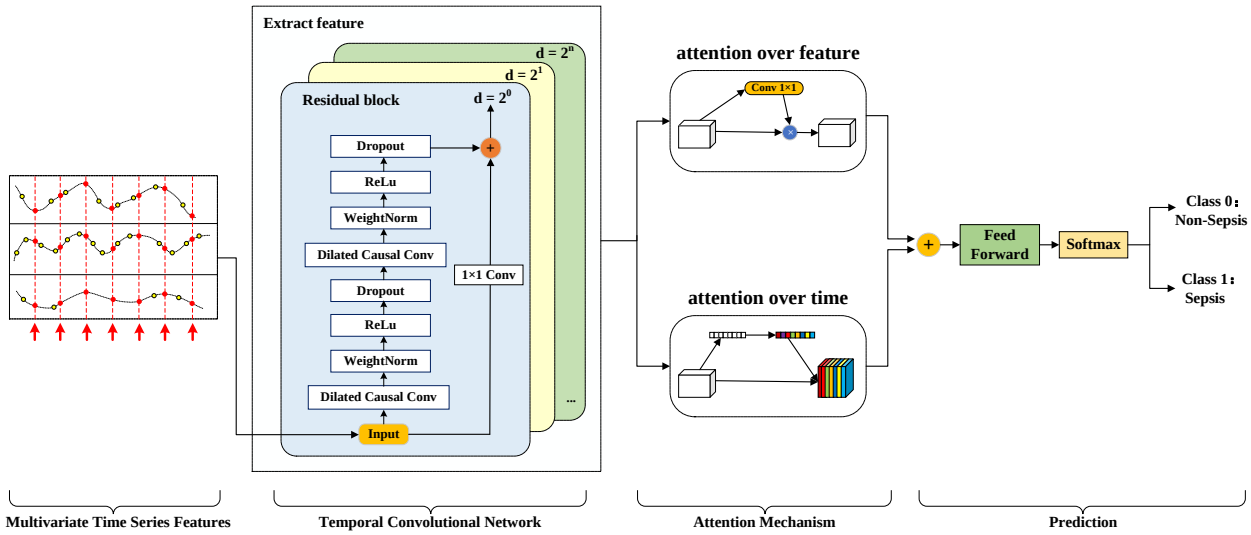

Figure 1. Overall framework of TCASP.

We propose the TCASP for sepsis diagnosis prediction, as shown in Figure 1. The model comprises four parts: a multivariate time-series feature vector Z input module, a temporal convolutional network module, an attention mechanism module, and a prediction module. The overall prediction process is as follows: first, the multivariate time series features of patients were obtained and input into the time convolutional network for feature learning; then, the extracted features are assigned different weights according to the feature channel attention and the time channel attention; finally, the probability of sepsis infection in this patient is derived using the softmax activation function.

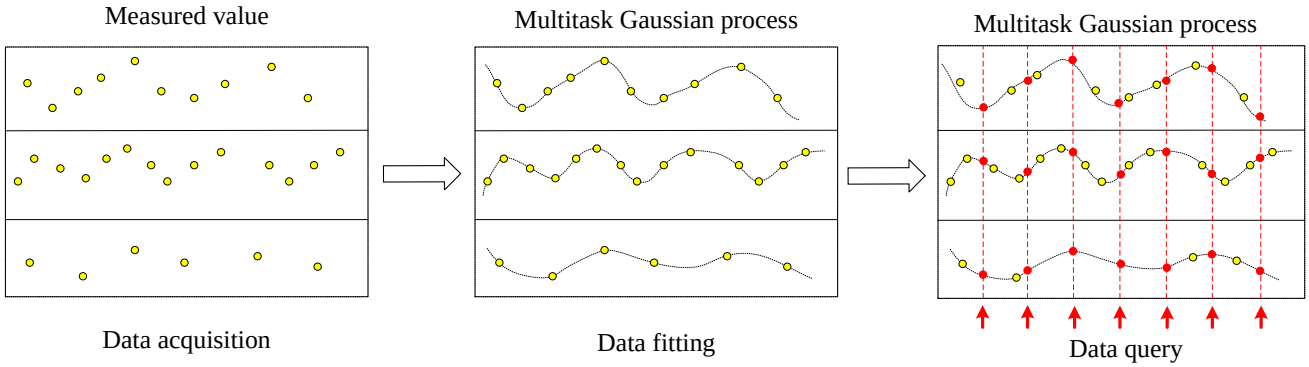

Figure 2. Multitask Gaussian process (MGP).

3.2. Multitask Gaussian process

Our proposed model uses an MGP [39] to preprocess the data. This method can handle irregular sampling frequencies, as shown in Figure 2. The MGP draws from the posterior distribution of a given observation at uniformly spaced grid times (hourly). More precisely, given irregularly sampled observations of i (value and time) $\{y_i, t_i\}$, for uniformly spaced query times x_i, the MGP plots a potential time series z_i based on the posterior distribution $P(z_i|y_i, t_i, x_i, \theta)$, as shown in Eq (3.1). The processed z_i is used as the input to the temporal convolutional network.

$$z_i \sim N(\mu(z_i), \Sigma(z_i); \theta) \tag{3.1}$$

Mean $\mu(z_i)$, and covariance $\Sigma(z_i)$ are, respectively:

$$\mu(z_i) = (K^D \otimes K^{X_i T_i})(K^D \otimes K^{T_i} + D \otimes I)^{-1} y_i \tag{3.2}$$

$$\Sigma(z_i) = (K^D \otimes K^{X_i}) - (K^D \otimes K^{X_i T_i})(K^D \otimes K^{T_i} + D \otimes I)^{-1}(K^D \otimes K^{T_i X_i}) \tag{3.3}$$

where $K^{X_i T_i}$ is the correlation matrix between the uniformly spaced data query x_i and the data fitting t_i, whereas denotes the correlation between x_i and itself. K^D is the task similarity kernel matrix whose $K^D_{d,d'}$ at position (d, d') denotes the similarity tasks (i.e., time-series channels) d and d'. Here, $\otimes$ denotes the tensor product, K^{T_i} denotes the specific T_i correlation matrix between all observed times encountered by patient i, and D is the diagonal matrix satisfying the pertask noise variance of $D_{dd} = \sigma_d^2$, with i is the unit matrix. The posterior mean $\mu(z_i)$ also depends on the observation y_i. We collect the parameters of the MGP in $\theta = \{K^D, \sigma_d^2|_{d=1}^D, l\}$, where l denotes the length scale of the kernel function.

3.3. Temporal convolutional network module

In the TCASP model, The multivariate time series feature vector Z_t of ICU patients was processed by MGP to obtain Z_i, which was input into the time convolutional network. The temporal convolutional network adds a dilated causal convolution based on a one-dimensional convolution. The causal convolution ensures that the output of the model at time t depends only on the information at t and the

moments before t and does not depend on the information after time t to prevent future information leakage. The dilated causal convolution is illustrated in Figure 3.

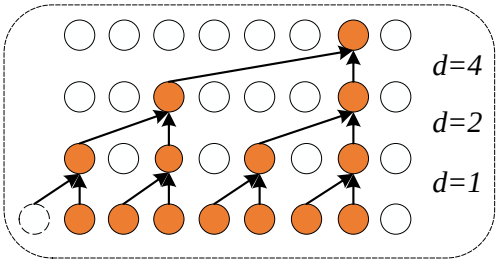

Figure 3. Dilated causal convolution.

For one-dimensional input sequence $x \in \mathbb{R}^n$ and a filter $f : \{0, 1, \cdots, k-1\} \rightarrow \mathbb{R}$, the dilated convolution operation F on element s of the sequence is defined as

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k} f(i) \cdot x_{s-d \cdot i} \tag{3.4}$$

where d is the expansion factor, k is the filter size, $x_{s-d \cdot i}$ represents the input sequence data, and input the sequence data $\{x_0, x_1, \cdots, x_t\}$ that need to be learned into the cause-and-effect extended convolution learning feature F_{tr}, and it can be seen from Figure 3 that when the number of neural network layers is large enough, the temporal convolutional network can learn all the input feature data and produce highly accurate prediction results.

Meanwhile, in order to avoid network degradation in the deep model, the temporal convolutional network adds residual connections between the convolutional layers. The output result of conversion F and the input x of this module are added together as the output result $F_{tr} = Activation(x + F(x))$ of the residual block, and the feature F_{tr} learned by the residual block is corrected.

3.4. Attention mechanism module

3.4.1. Feature channel attention

Medical datasets contain monitoring data of multiple patient physiological features that reflect changes in patient physiological features, but not all physiological features extracted by temporal convolutional networks are associated with sepsis. Therefore, this study proposes using feature channel attention to address this issue.

If the transformation in the temporal convolutional network is $U = F_{tr}(X)$, F_{tr} is denoted as a feature, if X is an input, then $X \in \mathbb{R}^{T \times N}$; if X is the output result of the intermediate layer, then $X \in \mathbb{R}^{T \times N \times C'}$. To facilitate the model description, this study used the output result $X \in \mathbb{R}^{T \times N \times C'}$ of the middle layer, and the operation of X as an input was similar to its operation as an output of the middle layer. $U \in \mathbb{R}^{T \times N \times C'}$, where T is the sequence length, N is the number of vital sign measurement terms, C' and C and are the number of channels.

First, 1×1 convolution is used in the original convolution result U to obtain the life feature channel attention m, as shown in Eq (3.5).

$$m = V_{1\times1} * U \tag{3.5}$$

where $m \in \mathbb{R}^{T\times N}$ is the learned feature attention, $V_{1\times1}$ is a convolution kernel, $U \in \mathbb{R}^{T\times N\times C'}$ is a transform in TCN, and $*$ is a convolution operation.

The feature channel attention and convolution results are then multiplied simultaneously as the result $\bar{x}_c$ output , as shown in Eq (3.6):

$$\bar{x}_c = m \cdot u_c \tag{3.6}$$

The size of the feature channel attention matrix is similar to that of the original convolution result, and the feature channel attention matrix is multiplied by the results under different channels as the result output.

Where $\bar{X} = \{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_c\}$, u_c is the convolution result U under the cth channel.

3.4.2. Time channel attention

Considering the interrelationship between channels, Hu et al. [40] introduced a novel structural unit- squeeze-and-excitation (SE) module in a CNN. The time channel attention proposed in this study is inspired by this approach. The ordinary channel attention mechanism is applied to the time interval, and the time channel attention is denoted as the time-SE module.

For narrative convenience, this subsection uses X as the output of the intermediate layer and X as the input for the analogous case. $V = \{v_1, v_2, \cdots, v_c\}$ denotes the convolution kernel used in the convolution, where v_c represents the parameters of the cth convolution kernel. The convolution result is denoted as $U = \{u_1, u_2, \cdots, u_c\}$, and the convolution result of the cth convolution kernel is.

$$u_c = v_c * X \tag{3.7}$$

From Eq (3.7), it is clear that the transformations in the temporal convolutional network alone cannot "focus" on the temporal information in the data that is relevant to the prediction task. In this study, we solve this problem using the temporal channel attention time-SE module.

The time-SE module, shown in Figure 4, is built in two steps: first, the squeeze module is built to extract the global information on the time "channel"; second, the excitation module calculates and reassigns the attention mechanism to the corresponding time interval based on the global information provided by squeeze.

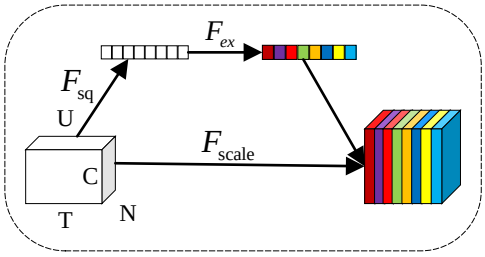

Figure 4. Causal dilation convolution.

Squeeze module: This module compresses the information of the convolution result of a temporal convolutional network in a time interval into a time descriptor. The time descriptor was implemented by considering the global average pooling operation to generate time-dependent statistics. Formally, a statistic $z \in \mathbb{R}^c$ is implemented by compressing U in the dimension $N \times C$. The tth element of z is computed by Eq (3.8):

$$z_t = F_{sq}(v_t) = \frac{1}{N \times C} \sum_{i=1}^{H} \sum_{i=1}^{H} v_t(i, j) \tag{3.8}$$

where z_t is the output of the squeeze module and, v_t is the convolution result at time t.

Excitation module: This module uses the squeeze module to obtain global information, calculate the attention of the time channel, and allocate different "importance" to different time intervals. To apply the gating mechanism to the excitation module, a formal calculation of this operation is shown in Eq (3.9):

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(w_2\sigma(W_1 z)) \tag{3.9}$$

where σ refers to the $ReLU$ activation function, $W_1 \in \mathbb{R}^{\frac{1}{r} \times T}$, and $W_2 \in \mathbb{R}^{\frac{1}{r} \times T}$, and s is the time channel attention mechanism. To limit the complexity of the model and thus enhance its generalization, two fully connected layers (FC) are added before the nonlinear transformation, that is, a descending layer with parameter W_1, a $ReLU$ layer, and an ascending layer with parameter W_2.

$$\tilde{x}_t = F_{scale}(v_t, s_t) = s_t \cdot v_t \tag{3.10}$$

where $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_t\}$, s_t is the time channel attention term assigned to time t and $v_t \in \mathbb{R}^{N \times C}$ is the result of convolving the data at time t. The time-SE module introduces dynamics in which the inputs are conditional (changes in time channel attention that occur with different inputs), which improves the ability of the model to distinguish the features of the data in time.

3.5. Prediction module

The feature channel attention m and time channel attention s_t are calculated separately on the convolution results, and then multiplied with the convolution result to obtain different attention output results $\bar{x}_c$ and $\tilde{x}_t$. The output results under the two-channel attention are summed to obtain the extracted features F_t , as shown in Eq (3.11):

$$F_t = \bar{x}_c + \tilde{x}_t \tag{3.11}$$

The features F_t are input into the feedforward neural network, and the probability of the onset of sepsis pt for the patient is derived using the softmax activation function, as shown in Eq (3.12):

$$p_t = softmax(F_t) \tag{3.12}$$

In this study, the cross-entropy loss function is used as the loss function L between the true value a_t and the predicted value p_t, and the minimization L is used to optimize the input of the patient's physiological features, network parameters and to predict the probability of sepsis onset in patients. The loss function L is calculated as shown in Eq (3.13).

$$L = -\sum_{i=1}^{N}[(a_t log p_t) + (1 - a_t) log(1 - p_t)] \tag{3.13}$$

3.6. TCASP prediction algorithm

The implementation of the temporal convolution attention model for sepsis diagnosis prediction is presented in Algorithm 1.

Algorithm 1 Algorithm of example

Input: Patient Information $H = \{h_1, h_2, \cdots, h_n\}$, $T = \{t_1, t_2, \cdots, t_j\}$, Number of layer L, Residual blocks B, Training hyperparameters

Output: The probability that the patient has sepsis p_t

for $h_t = \{Z_t | temp_k, heartrate_k, spo2_k, \cdots\}$ **do** // input physiological features into TCN
  $F(s) = (x *_d f)(s) = \sum_{i=0}^{k} f(i) \cdot x_{s-d \cdot i}$
  $F_{tr} = Activation(x + F(x))$
  for $\bar{x}_c = m \cdot u_c$ **do** // feature channel attention weighted
    $m = V_{1\times1} * U$
    $u_c = v_c * X$
  end for
  for $\tilde{x}_t = F_{scale}(v_t, s_t) = s_t \cdot v_t$ **do** // time channel attention weighted
    $z_t = F_{sq}(v_t) = \frac{1}{N\times C} \sum_{i=1}^{H} \sum_{i=1}^{H} v_t(i, j)$
    $s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(w_2 \sigma(W_1 z))$
    $v_t \in \mathbb{R}^{N\times C}$
  end for
end for
$F_t = \bar{x}_c + \tilde{x}_t$ // weighted sum of two sets of features
$p_t = softmax(F_t)$ // return to the incidence of sepsis in patients
Calculate the cross-entropy loss function L
Model optimization, parameter update
Return p_t // return the optimized result

4. Experiments and results

4.1. Experimental dataset

In this study, we used the EMRs dataset MIMIC-III [41] provided by Beth Israel Deaconess Medical Center, U.S.A. The MIMIC-III dataset contains information on multiple types of ICUs (surgical care units, medical care units, trauma surgical care units, neonatal care units, and cardiac care units) and includes vital signs, medications, laboratory measurements, medical orders, procedure codes, diagnosis codes, imaging reports, length of stay, and survival data. The longitudinal medical records used for this study collected data on over 58, 000 visits to 38, 645 adults and 7875 newborns seen at the medical center between 2001 and 2012. Institutional approval was obtained to use the deidentified data for

research purposes.

4.2. Data extraction and data preprocessing

4.2.1. Data extraction

Data were extracted from the MIMIC-III dataset according to the latest international consensus definition of sepsis, which is commonly used to assess the effectiveness of treatment in patients with suspected infection at risk of sepsis.

First, accurate and timely determination of the time of onset of sepsis is critical. SOFA (as shown in Table 1) is a scoring system for determining the degree of impairment of major organ function after a patient is admitted to the ICU ward, and a SOFA score ≥ 2 indicates that the patient is at risk of suspected sepsis infection. Therefore, in this study, the time of SOFA score ≥ 2 after the patient's admission to the ICU was determined as the onset of suspected sepsis infection.

Second, the following criteria were used for patient inclusion in the trial: (1) patients admitted to the ICU who were older than 18 years, with a high prevalence of sepsis; (2) ICU stay of over 24 h to ensure sufficient data for analysis, and (3) patients diagnosed with sepsis according to the international consensus definition. In addition, to ensure that nonseptic patients did not develop sepsis before ICU admission, the data center for nonseptic patients was required to have no ICD-9 codes associated with sepsis.

Finally, 24 dynamic features related to sepsis were selected for the experiment (as shown in Table 2). After screening the data, 17, 227 patients were finally selected, including 6769 patients with sepsis and 10, 458 patients without sepsis.

Table 1. Sequential organ failure assessment.

		Score				
System	**Variable**	**0**	**1**	**2**	**3**	**4**
Respiration	PaO2/FiO2, mmHg	≥ 400	< 400	< 300	< 200	< 100
	respiratory support				Yes	Yes
Coagulation	Platelets, 10^9/L	≥ 150	< 150	< 100	< 50	< 20
Liver	Bilirubin, μmol/L	< 20.5	≤ 34.1	≤ 102.5	≤ 205.1	> 205.2
Cardiovascular	MAP, mmHg	≥ 70	< 70			
	Dopamine, μg/(kg·min)			≤ 5	> 5	> 15
	dobutamine, μg/(kg·min)			any dose		
	epinephrine, μg/(kg·min)				≤ 0.1	> 0.1
	norepinephrine, μg/(kg·min)				≤ 0.1	> 0.1
Central nervous system	Glasgow Coma Scale score	15	13–14	10–12	6–9	< 6
Renal	Creatinine, μmol/L	< 110	≤ 176	≤ 308	≤ 442	> 442
	Urine output, ml/d				≤ 500	≤ 200

Note: (1) Daily assessment should take the worst daily value; (2) Higher scores indicate a worse prognosis.

4.2.2. Data preprocessing

The data obtained after extraction from the MIMIC-III dataset cannot be used directly for sepsis prediction, and further processing of the experimental data is required.

(1) Filling in missing data values: The typical solutions for missing data are deletion and mean interpolation; however, considering that medical data have sampling irregularity and uncertainty, there are certain problems in solving missing medical data using these methods. Therefore, the experiments used MGP to solve the missing data problem. It generates a continuous function to fit irregular samples by modeling the data covariance. Because it can easily account for uncertainty and does not require uniformly sampled data, it is the ideal solution for handling irregularly sampled medical time series.

(2) Data normalization: The prediction results are affected by the units of measurement of each feature value, for example, $90 \leq$ systolic blood pressure ≤ 140, $60 \leq$ diastolic blood pressure ≤ 90, $16 \leq$ respiratory rate ≤ 20. Normalization can limit the data to a certain range, thus eliminating the adverse effects caused by odd sample data. The normalization of the data accelerates the gradient descent to obtain the optimal solution and potentially improve accuracy.

Table 2. Dynamic features list of sepsis.

Vital Features	**Lab Features**	
Mean blood pressure	Platelet	Bicarbonate
Respiratory rate	Potassium	Creatinine
Heart rate	Pulse transit time	Chloride
Pulse oximetry	International normalized ratio	Glucose
Systolic blood pressure	Prothrombin time	Hematocrit
Diastolic blood pressure	Sodium	Hemoglobin
Body temperature	Blood urea nitrogen	Lactate
	White blood cells	Magnesium
		Blood gas pH

4.3. Benchmark model and experimental setup

To verify the effectiveness of the proposed TCASP model, five representative models were selected for comparison experiments:

Logistic regression [42] is a generalized linear regression analysis model that classifies data based on the available data by creating a regression formula for the classification boundary line.

Insight [43] is an early warning algorithm that has been developed for the early detection of septic infections in clinical settings.

RNN [44] is a type of neural network with short-term memory capacity and are used to process sequential data.

LSTM [45] network is a type of recurrent neural network after improvement, which can solve long-distance dependence.

TCN [46] is a network structure capable of processing time-series data and outperforming traditional neural networks under certain conditions.

In this study, we used TensorFlow 2.0 to implement the TCASP model. All experiments were conducted on a GPU server with 192G RAM, RTX6000 graphics card, and Ubuntu Server 18.04. The

dataset was divided into training, validation, and test sets at a ratio of 8 : 1 : 1. The experimental parameters of the TCASP model are listed in Table 3.

Table 3. Experimental parameters.

Parameter Description	Value
Learning rate	0.0005
Batch size	128
Number of epochs	100
Dropout rate	0.05
kernel size	2–6
Residual blocks	2–12
Hidden layers	10–55
L2 regularisation	0–250

4.4. Evaluation metrics

The sepsis prediction problem investigated in this study is essentially a binary classification learning problem. There are two main types of evaluation metrics for binary classification learning problems: AUROC and AUPRC. The sepsis prediction aims to identify a patient's risk of infection as early as possible, and the larger the area of the AUROC and AUPRC, the better the prediction effect. The indicators are briefly described as follows:

(1) AUROC: This metric is a randomly selected pair of samples (positive and negative samples) using a classifier that has been trained to predict a probability value for the positive and negative samples, respectively. The probability that the probability of the positive sample is greater than the probability of the negative sample is the AUROC, and the formula is

$$AUC = \frac{\sum_{ins_i \in positiveclass} rank_{ins_i} - \frac{M \times (M+1)}{2}}{M \times N} \tag{4.1}$$

where $\sum_{ins_i \in positiveclass}$ denotes the ordinal numbers of positive samples, $rank_{ins_i}$ denotes the number of the ith sample, and M, N denotes the number of positive and negative samples.

(2) AUPRC: The PRC curve is the line formed by the points of accuracy and recall, and is plotted with accuracy as the vertical axis and recall as the horizontal axis to obtain the recall-accuracy curve, referred to as the "PR curve".

In addition to the two evaluation metrics mentioned earlier, decision curve analysis (DCA) and clinical impact curve (CIC), which are widely used in clinical analysis, are used in this study for comparison of logistic, Insight, RNN, LSTM, TCN and TCASP models. The metrics are briefly described below.

(1) DCA is a simple method for evaluating clinical predictive models, diagnostic tests, and molecular markers. While traditional diagnostic assays, such as sensitivity, specificity, and area under the ROC curve, only measure the diagnostic accuracy of predictive models and fail to consider the clinical utility of a particular model, the DCA can integrate patient or decision-maker preferences into the analysis. This concept meets the practical needs of clinical decision making.

$$Nb = \frac{tpc}{n} - \frac{fpc}{n}\left(\frac{p_t}{1-p_t}\right) \tag{4.2}$$

where Nb is the net benefit, tpc is the true positive count, fpc is the false positive count, n is the total number, and p_t is the threshold probability.

(2) The CIC is a further development of the DCA algorithm, which is used in clinical settings to assess the predictive accuracy of predictive models.

4.5. Analysis of experimental results

This study aims to identify patients at risk of sepsis infection as early as possible and to minimize mortality in patients with sepsis. Therefore, the diagnostic accuracy of the TCASP prediction model is particularly important. Figures 5 shows the comparison of AURPC and AUPRC values for sepsis onset under different algorithms.

(1) The analysis of the experimental results in Figures 5 shows that within 6 hours before sepsis onset, the AUROC and AUPRC indexes of the TCASP model continue to rise, indicating that the probability of sepsis infection of patients increased continuously, with the highest values reaching 86.9% and 63.9%, respectively, which are far higher than those of other prediction models. This indicates that TCASP is the most suitable model for the early prediction of the first infection of sepsis in ICU patients. (2) As shown in the time complexity comparison in Table 4, where n is the number of training samples, t is the sequence length, and d is the vector dimension, the time complexity of the TCASP model in this study is $O(n * d^2 + t * d^2)$. The time complexity of logistics is the lowest among the compared methods, and the time complexity of TCASP is slightly higher than that of TCN. Therefore, there is a need for continuous optimization in terms of time complexity to improve the prediction performance of the model.

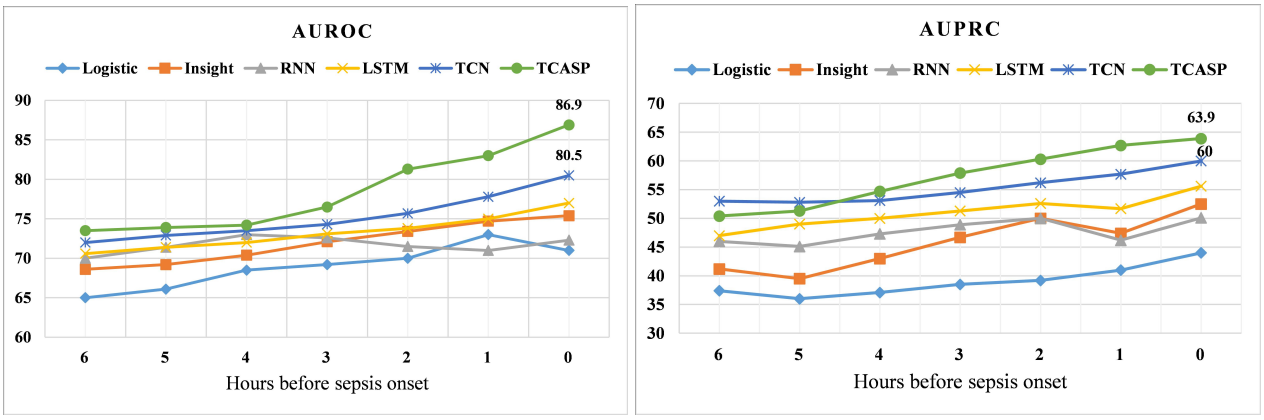

Figure 5. Comparison of AUROC and AUPRC values for sepsis onset under different algorithms.

Table 4. Time complexity comparison.

Algorithm	Logistic	Insight	RNN	LSTM	TCN	TCASP
Time Complexity	$O(n * d)$	$O(n * d^2)$	$O(n * d + t * d)$	$O(n * d^2 + t * d)$	$O(n * d^2 + t * d)$	$O(n * d^2 + t * d^2)$

To verify the effectiveness of MGP in the early diagnosis and prediction of sepsis, we compared five commonly used methods for handling missing values. According to the experimental results shown in Table 5, using MGP for missing value imputation yielded excellent results for the subsequent TCASP model in sepsis early diagnosis and prediction.This is because MGP can utilize the correlations between multiple tasks (i.e., different features) to impute missing values. By considering information from multiple features, the imputed values can more accurately reflect the actual condition of the patients, providing more comprehensive data. Single-task Gaussian process is also a flexible and effective method for handling missing values, but it is only suitable for imputing missing values in a single feature and cannot fully leverage the correlations between features. Nearest neighbor interpolation method can fill in missing values based on the nearest neighboring data points to the missing values, but it may not be robust enough in the presence of noise or outliers. Mean imputation and median imputation are simple and common methods that fill in missing values using the mean or median of the entire feature. These methods can preserve the overall statistical characteristics of the data to some extent but may overlook the correlations between features. Deleting missing values is the simplest approach, but it may result in a reduction in the amount of data and introduce bias in predictions.In summary, this study chose to use MGP to handle missing values in multivariate time series features because it can better leverage the correlations between features, providing more accurate and comprehensive data.

Table 5. Comparison of missing value handling methods.

Model	Missing Value Handling	AUROC	AUPRC
TCASP	Delete Missing Values	70.5 ± 2.8	50.4 ± 3.0
	Mean Imputation	73.1 ± 1.0	53.9 ± 1.7
	Median Imputation	76.8 ± 1.7	56.3 ± 0.9
	Nearest Neighbor Interpolation	79.4 ± 2.6	58.6 ± 1.5
	Single-task Gaussian Process	81.2 ± 1.2	60.1 ± 1.0
	Multi-task Gaussian Process	**86.9 ± 0.6**	**63.9 ± 1.3**

4.6. Analysis of ablation study results

In our model, we use an attention mechanism to enhance the selection of multivariate time-series features in the TCASP model, considering the specificity of medical datasets, which helps improve the prediction performance of the entire model. We conducted ablation experiments to demonstrate the effectiveness of the attention mechanism in the model for selecting important features and temporal information. The TCASP model consists of three parts: temporal convolutional network, feature channel attention (α), and time channel attention (β), and the entire TCASP model is split with the other three parts for comparison of the ablation experiments, and the results are shown in Tables 6 and 7.

Table 6. Area under the ROC curve for the ablation experiment.

Hours before sepsis onset	TCASP w/o α	TCASP w/o β	TCASP w/o attention	TCASP
6 h	65.2 ± 2.1	61.1 ± 1.9	57.4 ± 2.6	**73.5 ± 0.6**
5 h	67.5 ± 1.8	64.2 ± 1.0	60.9 ± 1.7	**73.9 ± 1.2**
4 h	69.8 ± 0.9	66.0 ± 0.4	62.8 ± 1.2	**74.2 ± 0.4**
3 h	71.2 ± 1.2	68.3 ± 0.7	64.3 ± 1.0	**76.5 ± 1.0**
2 h	73.4 ± 0.4	71.2 ± 0.9	69,4 ± 0.8	**81.3 ± 0.3**
1 h	76.1 ± 0.6	74.9 ± 0.5	72.6 ± 0.5	**83.0 ± 0.5**
0 h	79.5 ± 0.4	77.4 ± 0.7	74.2 ± 0.6	**86.9 ± 0.6**

Table 7. Area under the precision-recall curve for the ablation experiment.

Hours before sepsis onset	TCASP w/o α	TCASP w/o β	TCASP w/o attention	TCASP
6 h	48.5 ± 1.8	47.7 ± 1.0	45.7 ± 1.5	**50.4 ± 2.0**
5 h	49.3 ± 1.6	48.4 ± 1.5	47.0 ± 1.3	**51.3 ± 1.2**
4 h	50.1 ± 1.1	50.2 ± 1.2	49.3 ± 0.9	**54.7 ± 1.1**
3 h	51.6 ± 1.5	51.3 ± 1.3	50.6 ± 1.7	**57.9 ± 1.0**
2 h	53.7 ± 1.3	53.3 ± 1.7	51.1 ± 0.9	**60.3 ± 1.2**
1 h	55.1 ± 1.4	54.7 ± 1.8	52.0 ± 1.0	**62.7 ± 0.7**
0 h	59.9 ± 1.2	56.7 ± 1.2	54.4 ± 1.3	**63.9 ± 1.3**

To verify the effect of the attention mechanism on the importance of multivariate time series features, we removed the entire attention mechanism module at the input layer of feature extraction. To verify the effect of time channel attention (β) on the importance of selecting temporal information, we removed the time channel attention in the attention mechanism module. To verify the effect of the feature channel attention (α) on the importance of selecting important features, we removed the feature channel attention in the attention mechanism module.

By analyzing the results of the ablation experiments in Tables 6 and 7, we find that (1) the prediction performance is optimal when the model has the entire attention mechanism module, indicating that the attention mechanism significantly influences the model performance. (2) As the results of the ablation experiments indicate, the prediction performance of the model was significantly reduced by removing any part of the attention mechanism module. This is because the model does not select important features and temporal information more accurately after removing two-channel attention. (3) The ablation experiments show that as the number of patients in the EMRs increases, the patient data also increases exponentially. If important patient information is not extracted for the patient, it will bring a huge computational overhead to patient condition prediction and auxiliary diagnosis. The experiments in this study verify the effectiveness of the attention mechanism module in the TCASP model, which is beneficial for improving prediction performance.

4.7. Analysis of clinical trial results

In actual clinical diagnosis, we predict whether a patient has sepsis using the patient's physiological features, and no matter which feature value is selected as the threshold, we will encounter false positives and negatives, that is, misdiagnosis and underdiagnosis, sometimes avoiding false positives will

benefit the patient the most, and sometimes it is more desirable to avoid false negatives, and to attempt to avoid these two situations, decision curves are commonly used. Figure 6 shows the decision curve analysis of the six prediction models. The x-axis represents the threshold probability of critical care outcomes, and the y-axis represents the net benefit.

As shown in Figure 6, we can see that the TCASP model has the highest net benefit, while the Logistic regression model has the lowest net benefit. This is because the Logistic regression model has stricter requirements for features, needing them to satisfy the assumptions of linearity and independence. If there are non-linear features or highly correlated features in the sepsis prediction task, the Logistic regression model may not effectively utilize this information, thereby affecting the performance of net benefit. At the same time, as the threshold probability increases, the net benefit of the TCN model gradually decreases. This is due to the local nature of convolutional operations, which may pose challenges for TCN in handling long-term dependencies. In the sepsis prediction task, long-term temporal dependencies may be crucial for accurate predictions. The TCASP model effectively addresses both of these issues by combining temporal and feature attention mechanisms, achieving the best net benefit in the DCA curve. If the risk of a patient hi suffering from sepsis is recorded as p_i, and when p_i reaches the corresponding threshold, it is regarded to be positive, which means that the doctor can refer to the DCA index to take corresponding treatment measures in advance for patients suspected to be infected with sepsis; this will not only predict the probability that the patient may suffer from sepsis so that the patient's condition can be controlled early but also reduce the probability of misdiagnosis and avoid causing double losses to the patient's mental and property.

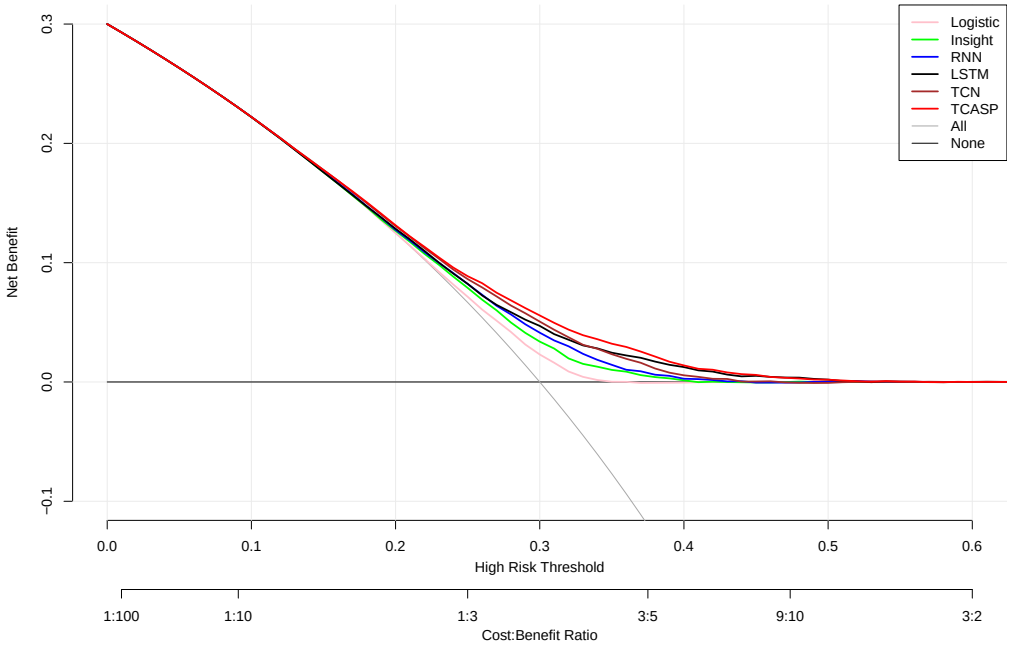

Figure 6. Decision curve analysis of six prediction models.

The CIC visualization of the TCASP model is shown in Figure 7. In the CIC, the x-axis indicates the threshold probability of critical care outcomes, and the y-axis indicates the number of high risk

individuals. The red curve indicates the number of people classified as positive (high risk population) by the TCASP model at each threshold, and the blue curve indicates the number of true positives (actual occurrence of the high risk population) at each threshold.

As shown in Figure 7, according to the CIC analysis results, when the threshold exceeds 50%, the TCASP model identifies a number of individuals as high-risk sepsis patients that highly matches the actual number of high-risk sepsis patients. This indicates that the TCASP model has high predictive accuracy at high thresholds, meaning that the model tends to classify individuals with higher probabilities as high-risk patients. At the same time, the model's positive predictions match the actual number of positives, which means that at high thresholds, the model can effectively identify true high-risk patients, thus providing good clinical utility. This is crucial for the early diagnosis and intervention of sepsis, as it helps doctors better identify high-risk patients and take appropriate treatment measures promptly. However, it is important to note that for lower threshold probabilities, the model's predictions may have some errors, meaning that the model may miss some true high-risk individuals. Therefore, in practical applications, doctors and clinical decision-makers need to balance the trade-off between accuracy and missed diagnosis rate and choose an appropriate threshold based on specific circumstances.

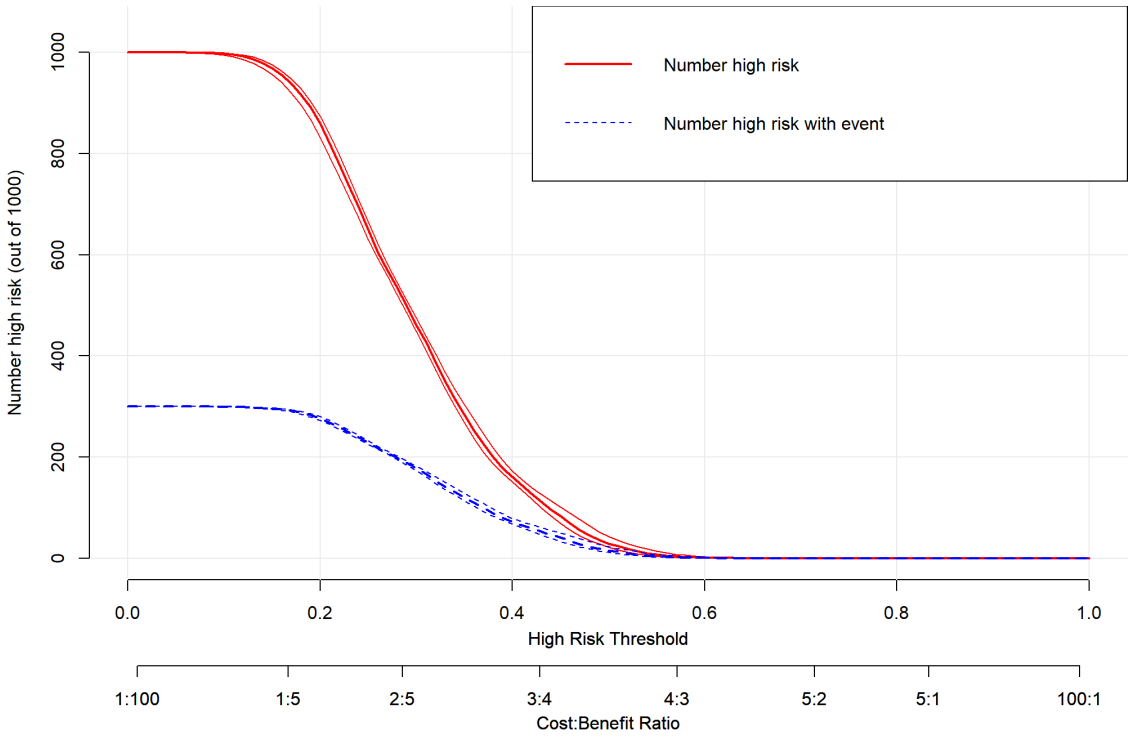

Figure 7. Clinical impact curve of the TCASP model.

5. Conclusions

We proposed the TCASP model comprehensively considered the uncertainty of EMRs and used the multitask Gaussian process to tackle the missing data. We introduced a dual channel attention mechanism, which enables the model to "independently" select relevant features, and solve the problem of difficulty in focusing on important features and time information owing to the huge and complex EMRs. The TCASP model can focus on the time-series information and predict the time of first sepsis infection in ICU patients.

However, the TCASP model also has some shortcomings: (1) Although introducing the time channel attention mechanism in the model can focus on time information, which can improve the effect of feature learning, it also improves the time complexity. (2) The experimental data in this study came only from an internationally open dataset. In this dataset, the proportion of male patients was relatively high, and the number of white patients was relatively high, which may have led to a potential demographic bias. (3) Our dataset analysis may be insufficient, leading to some key feature variables not being included in the TCASP model.

In future studies, we aim to further improve the model to enable it to identify high risk sepsis patients more comprehensively, accurately, and promptly. We aim to further improve the TCASP model, reduce the time complexity, provide technical support for clinical assistant diagnosis, personalized diagnosis, and treatment of sepsis, and improve the level of smart medical services.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments. This research was funded by the Northwest Normal University Major Research Project Incubation Program, China (No. NWNU-LKZD2021-06).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. M. S. Hari, G. S. Phillips, M. L. Levy, C. W. Seymour, V. X. Liu, C. S. Deutschman, et al., Developing a new definition and assessing new clinical criteria for septic shock: For the third international consensus definitions for sepsis and septic shock (sepsis-3), *JAMA*, **315** (2016), 775–787. https://doi.org/10.1001/jama.2016.0289
2. C. Fleischmann-Struzek, D. M. Goldfarb, P. Schlattmann, L. J. Schlapbach, K. Reinhart, N. Kissoon, The global burden of paediatric and neonatal sepsis: A systematic review, *Lancet Respir. Med.*, **6** (2018), 223–230. https://doi.org/10.1016/S2213-2600(18)30063-8
3. C. Fleischmann, D. O. Thomas-Rueddel, M. Hartmann, C. S. Hartog, T. Welte, S. Heublein, et al., Hospital incidence and mortality rates of sepsis: An analysis of hospital episode (DRG) statistics in germany from 2007 to 2013, *Deutsch. Ärzteblatt Int.*, **113** (2016), 159. https://doi.org/10.3238%2Farztebl.2016.0159

4. S. M. Perman, M. Goyal, D. F. Gaieski, Initial emergency department diagnosis and management of adult patients with severe sepsis and septic shock, *Scand. J. Trauma Resusc. Emerg. Med.*, **20** (2012), 1–11. https://doi.org/10.1186/1757-7241-20-41
5. K. E. Rudd, S. C. Johnson, K. M. Agesa, K. A. Shackelford, D. Tsoi, D. R. Kievlan, et al., Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the global burden of disease study, *Lancet*, **395** (2020), 200–211. https://doi.org/10.1016/S0140-6736(19)32989-7
6. M. S. Rangel-Frausto, D. Pittet, M. Costigan, T. Hwang, C. S. Davis, R. P. Wenzel, The natural history of the systemic inflammatory response syndrome (SIRS): A prospective study, *Jama*, **273** (1995), 117–123. https://doi.org/10.1001/jama.1995.03520260039030
7. Á. Castellanos-Ortega, B. Suberviola, L. A. García-Astudillo, M. S. Holanda, F. Ortiz, J. Llorca, et al., Impact of the surviving sepsis campaign protocols on hospital length of stay and mortality in septic shock patients: Results of a three-year follow-up quasi-experimental study, *Criti. Care Med.*, **38** (2010), 1036–1043. https://doi.org/10.1097/CCM.0b013e3181d455b6
8. J. K. Sandhu, U. K. Lilhore, M. Poongodi, N. Kaur, S. S. Band, M. Hamdi, et al., Predicting the risk of heart failure based on clinical data, *Hum. Centric Comput. Inf. Sci.*, **12** (2022).
9. S. Thandapani, M. I. Mahaboob, C. Iwendi, D. Selvaraj, A. Dumka, M. Rashid, et al., IoMT with deep CNN: Ai-based intelligent support system for pandemic diseases, *Electronics*, **12** (2023), 424. https://doi.org/10.3390/electronics12020424
10. E. M. Onyema, S. Balasubaramanian, S. K. Suguna, C. Iwendi, B. S. Prasad, C. D. Edeh, Remote monitoring system using slow-fast deep convolution neural network model for identifying anti-social activities in surveillance applications, *Meas. Sensors*, **27** (2023), 100718. https://doi.org/10.1016/j.measen.2023.100718
11. H. J. Kam, H. Y. Kim, Learning representations for the early detection of sepsis with deep neural networks, *Comput. Biol. Med.*, **89** (2017), 248–255. https://doi.org/10.1016/j.compbiomed.2017.08.015
12. M. A. Reyna, C. Josef, S. Seyedi, R. Jeter, S. P. Shashikumar, M. B. Westover, et al., Early prediction of sepsis from clinical data: The physionet/computing in cardiology challenge 2019, in *2019 Computing in Cardiology (CinC)*, (2019), 1–4. https://doi.org/10.22489/CinC.2019.412
13. S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, T. G. Buchman, An interpretable machine learning model for accurate prediction of sepsis in the ICU, *Crit. Care Med.*, **46** (2018), 547. https://doi.org/10.1097%2FCCM.0000000000002936
14. E. Sheetrit, N. Nissim, D. Klimov, Y. Shahar, Temporal probabilistic profiles for sepsis prediction in the ICU, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (2019), 2961–2969. https://doi.org/10.1145/3292500.3330747
15. L. M. Fleuren, T. L. Klausch, C. L. Zwager, L. J. Schoonmade, T. Guo, L. F. Roggeveen, et al., Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy, *Intensive Care Med.*, **46** (2020), 383–400. https://doi.org/10.1007/s00134-019-05872-y
16. A. Wong, E. Otles, J. P. Donnelly, A. Krumm, J. McCullough, O. DeTroyer-Cooley, et al., External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients, *JAMA Int. Med.*, **181** (2021), 1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626

17. K. Rahmani, R. Thapa, P. Tsou, S. C. Chetty, G. Barnes, C. Lam, et al., Assessing the effects of data drift on the performance of machine learning models used in clinical sepsis prediction, *Int. J. Med. Inf.*, **173** (2023), 104930. https://doi.org/10.1016/j.ijmedinf.2022.104930

18. R. C. Bone, R. A. Balk, F. B. Cerra, R. P. Dellinger, A. M. Fein, W. A. Knaus, et al., Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis, *Chest*, **101** (1992), 1644–1655. https://doi.org/10.1378/chest.101.6.1644

19. J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, et al., *The sofa (sepsis-related organ failure assessment) score to describe organ dysfunction/failure: On behalf of the working group on sepsis-related problems of the european society of intensive care medicine (see contributors to the project in the appendix)*, 1996. Available from: http://pirasoa.iavante.es/pluginfile.php/4037/mod_label/intro/18.%20Vincent%201996.pdf.

20. C. Stenhouse, S. Coates, M. Tivey, P. Allsop, T. Parker, Prospective evaluation of a modified early warning score to aid earlier detection of patients developing critical illness on a general surgical ward, *Br. J. Anaesth.*, **84** (2000), 663. https://doi.org/10.1093/bja/84.5.663

21. O. A. Usman, A. A. Usman, M. A. Ward, Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the emergency department, *Am. J. Emerg. Med.*, **37** (2019), 1490–1497. https://doi.org/10.1016/j.ajem.2018.10.058

22. A. E. Johnson, J. Aboab, J. D. Raffa, T. J. Pollard, R. O. Deliberato, L. A. Celi, et al., A comparative analysis of sepsis identification methods in an electronic database, *Crit. Care Med.*, **46** (2018), 494. https://doi.org/10.1097%2FCCM.0000000000002965

23. S. Van der Woude, F. Van Doormaal, B. Hutten, F. Nellen, F. Holleman, Classifying sepsis patients in the emergency department using SIRS, qSOFA or MEWS, *Neth. J. Med.*, **76** (2018), 158–166.

24. K. E. Henry, D. N. Hager, P. J. Pronovost, S. Saria, A targeted real-time early warning score (TREWScore) for septic shock, *Sci. Transl. Med.*, **7** (2015), 299ra122. https://doi.org/10.1126/scitranslmed.aab3719

25. S. Horng, D. A. Sontag, Y. Halpern, Y. Jernite, N. I. Shapiro, L. A. Nathanson, Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning, *PloS One*, **12** (2017), e0174708. https://doi.org/10.1371/journal.pone.0174708

26. S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, T. G. Buchman, An interpretable machine learning model for accurate prediction of sepsis in the ICU, *Crit. Care Med.*, **46** (2018), 547. https://doi.org/10.1097%2FCCM.0000000000002936

27. N. Wu, B. Green, X. Ben, S. O'Banion, Deep transformer models for time series forecasting: The influenza prevalence case, preprint, arXiv:2001.08317. https://doi.org/10.48550/arXiv.2001.08317

28. H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, et al., Informer: Beyond efficient transformer for long sequence time-series forecasting, **35** (2021), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325

29. S. S. Rangapuram, M. W. Seeger, J. Gasthaus, L. Stella, Y. Wang, T. Januschowski, Deep state space models for time series forecasting, *Adv. Neural Inf. Process. Syst.*, **2018** (2018), 31.

30. C. Lin, Y. Zhang, J. Ivy, M. Capan, R. Arnold, J. M. Huddleston, et al., Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-LSTM, **2018** (2018), 219–228. https://doi.org/10.1109/ICHI.2018.00032

31. S. Baral, A. Alsadoon, P. Prasad, S. Al Aloussi, O. H. Alsadoon, A novel solution of using deep learning for early prediction cardiac arrest in sepsis patient: Enhanced bidirectional long short-term memory (LSTM), *Multimedia Tools Appl.*, **80** (2021), 32639–32664. https://doi.org/10.1007/s11042-021-11176-5

32. A. Rafiei, A. Rezaee, F. Hajati, S. Gheisari, M. Golzan, SSP: Early prediction of sepsis using fully connected lstm-cnn model, *Comput. Biol. Med.*, **128** (2021), 104110. https://doi.org/10.1016/j.compbiomed.2020.104110

33. M. Saqib, Y. Sha, M. D. Wang, Early prediction of sepsis in EMR records using traditional ML techniques and deep learning LSTM networks, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (2018), 4038–4041. https://doi.org/10.1109/EMBC.2018.8513254

34. D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, preprint, arXiv:1409.0473. https://doi.org/10.48550/arXiv.1409.0473

35. E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, W. Stewart, Retain: An interpretable predictive model for healthcare using reverse time attention mechanism, *Adv. Neural Inf. Process. Syst.*, **2016** (2016), 29.

36. M. Usama, B. Ahmad, W. Xiao, M. S. Hossain, G. Muhammad, Self-attention based recurrent convolutional neural network for disease prediction using healthcare data, *Comput. Methods Programs Biomed.*, **190** (2020), 05191. https://doi.org/10.1016/j.cmpb.2019.105191

37. W. Lan, X. Wu, Q. Chen, W. Peng, J. Wang, Y. P. Chen, GANLDA: Graph attention network for lncrna-disease associations prediction, *Neurocomputing*, **469** (2022), 384–393. https://doi.org/10.1016/j.neucom.2020.09.094

38. L. Lin, B. Xu, W. Wu, T. W. Richardson, E. A. Bernal, Medical time series classification with hierarchical attention-based temporal convolutional networks: A case study of myotonic dystrophy diagnosis, in *CVPR Workshops*, (2019), 83–86.

39. E. V. Bonilla, K. Chai, C. Williams, Multi-task gaussian process prediction, *Adv. Neural Inf. Proces. Syst.*, **2007** (2007), 20.

40. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

41. A. E. Johnson, T. J. Pollard, L. Shen, L. W. H. Lehman, M. Feng, M. Ghassemi, et al., MIMIC-III, a freely accessible critical care database, *Sci. Data*, **3** (2016), 1–9. https://doi.org/10.1038/sdata.2016.35

42. F. S. De Menezes, G. R. Liska, M. A. Cirillo, M. J. Vivanco, Data classification with binary response through the boosting algorithm and logistic regression, *Exp. Syst. Appl.*, **69** (2017), 62–73. https://doi.org/10.1016/j.eswa.2016.08.014

43. J. S. Calvert, D. A. Price, U. K. Chettipally, C. W. Barton, M. D. Feldman, J. L. Hoffman, et al., A computational approach to early sepsis detection, *Comput. Biol. Med.*, **74** (2016), 69–73. https://doi.org/10.1016/j.compbiomed.2016.05.003

44. J. Futoma, S. Hariharan, K. Heller, M. Sendak, N. Brajer, M. Clement, et al., An improved multi-output gaussian process rnn with real-time validation for early sepsis detection, in *Machine Learning for Healthcare Conference*, (2017), 243–254.

45. K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, J. Schmidhuber, LSTM: A search space odyssey, *IEEE Trans. Neural Networks Learn. Syst.*, **28** (2016), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

46. M. Moor, M. Horn, B. Rieck, D. Roqueiro, K. Borgwardt, Early recognition of sepsis with gaussian process temporal convolutional networks and dynamic time warping, in *Machine Learning for Healthcare Conference*, (2019), 2–26.

AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)