International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-024-00568-z

REGULAR PAPER

An attention-based bidirectional LSTM-CNN architecture for the early prediction of sepsis

Pronaya Prosun Das[1,2] · Lena Wiese[1,2] · Marcel Mast[3] · Julia Böhnke[4] · Antje Wulff[3,6] · Michael Marschollek[3] · Louisa Bode[3] · Henning Rathert[5] · Thomas Jack[5] · Sven Schamer[5] · Philipp Beerbaum[5] · Nicole Rübsamen[4] · Andrè Karch[4] · Christian Groszweski-Anders[7] · Andreas Haller[7] · Torsten Frank[7]

Received: 30 November 2023 / Accepted: 18 May 2024

Abstract
Sepsis is a severe and expensive medical emergency that requires prompt identification in order to improve patient mortality. The objective of our research is to develop an attention-based bidirectional LSTM-CNN (AT-BiLSTM-CNN) hybrid architecture for the early prediction of sepsis using electronic health records (EHRs) obtained from intensive care units (ICUs). We combine attention mechanism, bidirectional long short-term memory (BiLSTM) and convolutional neural network (CNN) to analyse clinical time series data, aiming to enhance prediction accuracy. The effectiveness of our model is measured using metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic (AUROC), utilising data from the 2019 PhysioNet Challenge. Upon assessing the performance of the AT-BiLSTM-CNN model throughout prediction windows of 4, 8, and 12 h, we observed its exceptional performance in comparison with existing leading techniques. It achieved average AUROCs of 0.88, 0.85, and 0.84 for the predictions made 4, 8, and 12 h before sepsis onset, respectively. This research contributes significantly to the development of smart clinical support systems, potentially offering lifesaving interventions for septic patients at critical moments.

Keywords Attention · BiLSTM · Deep learning · Multivariate time series · Sepsis · EHRs

1 Introduction

Annually, a substantial proportion of individuals worldwide are admitted to intensive care units (ICUs) [20]. These units collect and store comprehensive patient data, such as vital signs, test outcomes, and personal details [26]. Doctors in these demanding settings encounter the task of making critical decisions by analysing a wide range of complex clinical and physiological information within strict time constraints and amidst ambiguity. AI-based technologies provide a solution to improve continuous patient monitoring and efficiently evaluate and comprehend various data sources, assisting in timely and targeted medical interventions [11]. AI is a more effective and preferable choice over traditional statistical methods when handling complex and extensive variables, such as the ones present in ICU data.

Sepsis, an acute syndrome triggered by the body's response to infection, is a prominent contributor to mortality worldwide and poses a substantial health obstacle. Sepsis, which was initially defined for adults in 1992, can impact people of all age groups and is associated with higher

✉ Pronaya Prosun Das
pronaya.prosun@stud.uni-frankfurt.de

Lena Wiese
lena.wiese@item.fraunhofer.de

1 Fraunhofer Institute for Toxicology and Experimental Medicine, Nikolai-Fuchs-Strasse 1, 30625 Hanover, Germany

2 Institute of Computer Science, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 1, 60629 Frankfurt am Main, Germany

3 Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover Medical School, Hannover, Germany

4 Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany

5 Department of Pediatric Cardiology and Intensive Care Medicine, Hanover Medical School, Hanover, Germany

6 Big Data in Medicine, Department of Health Services Research, School of Medicine and Health Sciences, Carl von Ossietzky University Oldenburg, Oldenburg, Germany

7 medisite GmbH, Hanover, Germany

Published online: 03 June 2024

mortality rates [4]. For example, in the USA, around 1.7 million adults receive a diagnosis of sepsis every year, resulting in over 270,000 fatalities annually [7]. Furthermore, it has the position of being the most costly ailment to manage, with US medical facilities incurring approximately $24 billion in expenses in 2013 [54]. If sepsis is not promptly and appropriately treated, it can progress to more serious conditions, such as severe sepsis or septic shock, which significantly heightens the likelihood of death [38]. Studies suggest that promptly identifying sepsis can result in timely intervention and markedly improved patient outcomes [43, 51]. Diagnosing sepsis in its early stages is a challenge due to the resemblance of its symptoms, such as fever, chills, fast breathing, and high heart rate, to those of other ailments. Furthermore, the ability to accurately forecast sepsis only a few minutes before its onset is not of significant clinical value. An optimal prediction model should ideally offer timely alerts and more robust indicators as the event approaches.

Electronic health records (EHRs) are computerised repositories of complete patient health information that have experienced a substantial surge in both quantity and diversity in recent times. The growth has facilitated the utilisation of machine learning and data mining approaches for the timely identification of sepsis. Clinically, sepsis can be identified using different screening tools such as the quick Sequential (Sepsis-Related) Organ Failure Assessment (qSOFA) [51], Systemic Inflammatory Response Syndrome (SIRS) [4], National Early Warning Score (NEWS) [52], and Modified Early Warning Score (MEWS) [53]. Nevertheless, these methods primarily focus on identifying preexisting symptoms rather than actively forecasting the start of sepsis, resulting in constraints on their diagnostic efficacy. Research has shown that the qSOFA score has limited sensitivity in identifying sepsis in prehospital settings and emergency departments [15, 56].

Lately, ICUs have implemented patient data management systems (PDMSs) to monitor patients and maintain clinical records. These sophisticated systems have the ability to directly display and examine all incoming data, which is essential for diagnosing conditions such as sepsis and Systemic Inflammatory Response Syndrome (SIRS). This progress presents significant possibilities for enhancing clinical diagnoses through data mining and improving patient outcomes using machine learning techniques [3, 21, 29].

Current studies have focused on utilising machine learning techniques to forecast sepsis in its early stages. Faisal et al. developed a logistic regression model called CARS, which aims to predict the likelihood of sepsis by utilising a patient's initial vital signs and blood test results. These data can typically be obtained within a few hours of emergency admission [17]. Mollura et al. constructed a bagged tree classifier by utilising data obtained from electrocardiograms and arterial blood pressure waveforms. According to their research, observing these waveforms can help in detecting sepsis within the first hour of a patient's admission to the ICU [40]. Desautels et al. developed a specialised machine learning algorithm called InSight, which uses vital signs, age, Glasgow Coma Score (GCS), and pulse oximetry as input variables [13]. The efficacy of this approach was demonstrated by obtaining a 57% accuracy and an area under the receiver operating characteristics (AUROC) curve of 0.74 in predicting the onset of sepsis four hours in advance. The InSight algorithm demonstrated superior performance compared to expert scoring systems in the prediction of sepsis in hospital environments. Horng et al. created a machine learning model that utilised a linear support vector machine. They demonstrated the benefit of including free text data in addition to vital signs and demographic information [22]. This method improves the provision of clinical guidance for sepsis in emergency departments. In 2018, Mao et al. validated the InSight algorithm, which identifies and forecasts three crucial sepsis indicators (SIRS, SOFA, and MEWS) [37]. The researchers utilised data from four distinct institutions and employed transfer learning techniques to evaluate its extensive applicability. The algorithm integrated six essential physiological indicators: respiratory rate, heart rate, peripheral capillary oxygen saturation, body temperature, diastolic blood pressure, and systolic blood pressure. The AISE algorithm, developed by Nemati et al., is specifically designed to accurately forecast the occurrence of sepsis at an early stage [41]. AISE integrates lower-resolution electronic health records (EHRs) with comprehensive data on blood pressure and heart rate. AISE has demonstrated a prediction accuracy of 63–67% and an AUROC score ranging from 0.83 to 0.85 when forecasting within a time frame of 4–12 h in advance. Kamaleswaran et al. established that by analysing physiomarkers, artificial intelligence has the capability to forecast the occurrence of severe sepsis in critically ill children with a lead time of up to 8 h [28]. Lyra et al. utilised an optimised random forest model to predict sepsis using imbalanced clinical data from ICUs, as a component of the PhysioNet Computing in Cardiology Challenge 2019 [6, 45].

In recent years, deep learning techniques have outperformed conventional models in terms of performance, demonstrating significant potential in the healthcare sector [39]. These deep learning models have the ability to independently learn from data representations, improving performance without relying on traditional feature extraction methods. Recurrent neural networks (RNNs) are commonly used in these models, particularly for forecasting results from multivariate time series data [8, 23]. Lauritsen et al. devised a sophisticated deep network model with the specific objective of predicting sepsis at an early stage [34]. The effectiveness of this model was evaluated by comparing it to a traditional multilayer feedforward neural network, using a diverse dataset collected from multiple centres. Their findings demonstrated

an AUROC of 0.84 for sepsis predictions made 3 h before the onset, and 0.79 for predictions made 10 h before the onset. This highlights the growing need for the development of autonomous systems that can assist in the early detection of diseases. Scherpf et al. utilised RNN to detect sepsis at an early stage, following the established benchmark set by Calvert et al. in their computational study. Their RNN is composed of two hidden layers of Gated Recurrent Units (GRUs), with each layer containing 40 units. Although the model demonstrated superior performance compared to the InSight algorithm in terms of the area under the receiver operating characteristics (AUROC), its practical applicability in a clinical setting is questionable due to its low specificity of only 47%. This low specificity leads to a significant number of false negatives.

Existing sepsis prediction models face two primary challenges: inadequate performance and a limited prediction window. In order to address these issues, we have devised an innovative approach for promptly and early identifying sepsis in patients in the ICU. The methodology we employ utilises an attention-based BiLSTM-CNN (AT-BiLSTM-CNN) algorithm, which is rooted in the principles of deep learning. This algorithm underwent extensive training, testing, and validation using a dataset consisting of 40,336 health records from ICU admissions. The findings of our study indicate that our AT-BiLSTM-CNN model outperforms existing models in terms of accuracy, as evidenced by an average AUROC of 0.88 for predictions made 4 h before the onset of sepsis. Moreover, it demonstrates improved adaptability and durability in managing data with low density and replacing absent values. Importantly, the AT-BiLSTM-CNN model can predict sepsis up to 12 h in advance.

The layout of the paper is arranged in the following manner: Section 2 presents the proposed methodologies along with the materials. The empirical evaluation is outlined in Sect. 3. Finally, Sect. 4 provides a conclusion summarising the findings of the research.

2 Material and methods

Figure 1 illustrates the layout of the framework. It processes input data, which includes vital signs, laboratory test results, and demographic information. Initially, the framework undergoes a preprocessing phase, where it fills in any missing values and removes unsuitable records. Subsequently, it employs a deep network block, featuring attention, bidirectional long short-term memory (LSTM), convolutional, and fully connected layers to predict the likelihood of sepsis development. The model then uses a predetermined reference threshold to determine if there is a suspicion of sepsis in the patient.

2.1 Dataset

Our training process utilised the 2019 PhysioNet Computing in Cardiology Challenge dataset [46], which includes around 40,000 electronic health records from ICU patients, either suspected of or confirmed with sepsis. These records were sourced from both Beth Israel Deaconess Medical Center and Emory University Hospital. Each record in this dataset presents a sequential timeline of clinical observations recorded hourly. They also feature a binary sepsis indicator for each time point, formatted as a single row. The dataset encompasses 40 different numerical and categorical clinical attributes, covering areas like patient demographics, vital signs, and lab results. A notable aspect of this dataset is the irregularity in data recording frequencies, attributed to the prevalence of missing values. These clinical attributes have been compiled over the last ten years by medical professionals, with all processes approved by the relevant Institutional Review Boards.

Each patient has been labelled as "sepsis" or "non-sepsis" by the dataset providers. This binary classification defines whether or not a patient meets the gold standard of sepsis according to the Sepsis-3 guidelines, i.e. a two-point change in the patient's Sequential Organ Failure Assessment (SOFA) score and clinical suspicion of infection. In the case of "sepsis", the onset time has also been provided. This onset time is determined by identifying a period where there's a suspected infection along with a SOFA score change of two or more points, spanning from 24 h before to 12 h after the initial suspicion of infection. The dataset also considers the earliest instance of either antibiotic administration or culture draw as an indicator of clinical suspicion of infection [51]. The dataset includes records of 2932 patients with sepsis and 37,404 without sepsis.

Our work incorporated vital signs such as heart rate, body temperature, diastolic and systolic blood pressure, oxygen saturation (O2Sat), and breathing rate. We also used laboratory test outcomes, including levels of serum glucose, white blood cell (Leukocyte) count, creatinine, platelet count, haemoglobin, haematocrit, and blood urea nitrogen (BUN). Factors like the patient's age and gender, the length of their stay in the ICU, and the duration between their hospital and ICU admissions were considered as demographic variables to aid in the early detection of sepsis.

2.2 Preprocessing

The framework begins with a crucial first step: preprocessing the data prior to its introduction into the deep network block. As shown in Fig. 2, this process involves several key actions. Initially, we addressed the issue of missing values by implementing imputation methods. Following this, moving averages for the vital signs are calculated to ensure

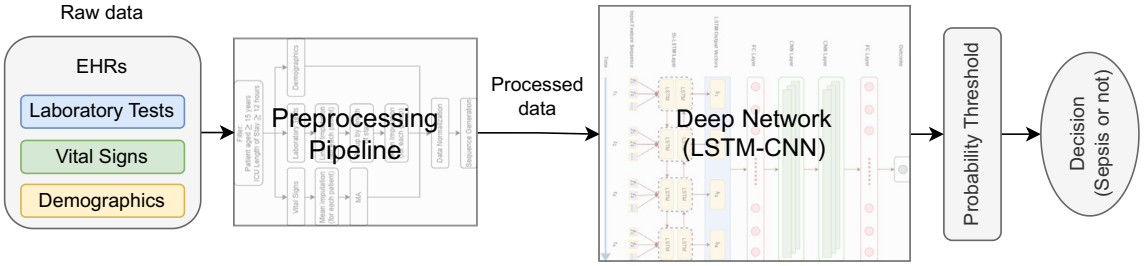

Fig. 1 Overall block diagram of our proposed approach. Preprocessing steps and deep network architecture are shown separately in Figs. 2 and 4, respectively

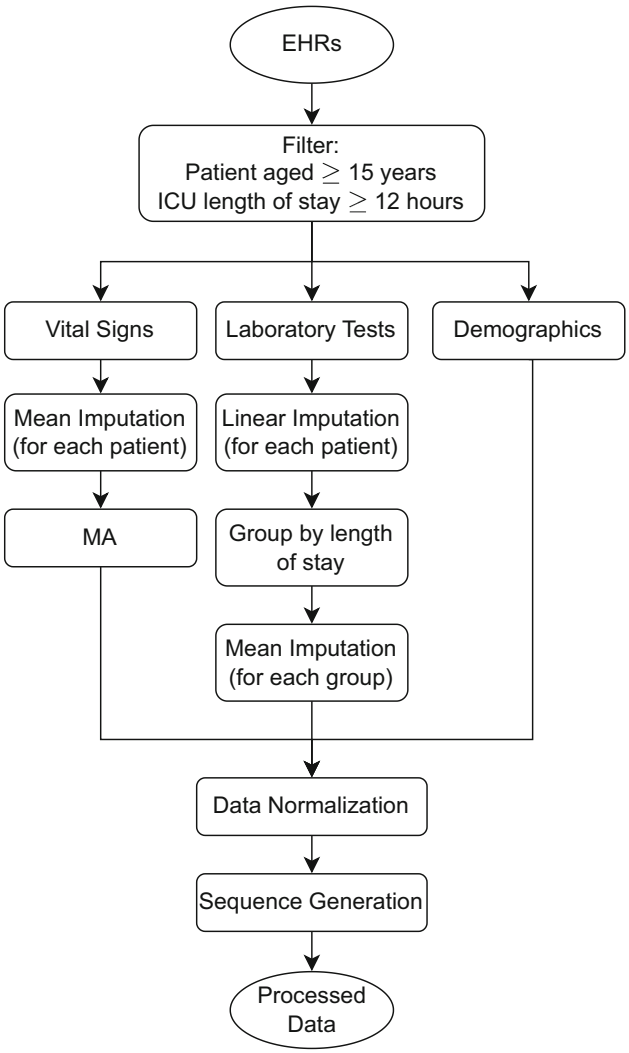

Fig. 2 Steps in preprocessing block

data consistency. The focus then shifted towards selecting patient records, specifically targeting those aged 15 years and above, in recognition of the differing sepsis criteria in children. Finally, we narrowed down the dataset further, opting to include only those records where the patient's ICU stay exceeded 12 h.

Table 1 An example imputation for leukocyte count

Hour	Real values	Imputed values
2		7.4
3	7.4	7.4
4		7.8
5		8.2
6	8.6	8.6
7		8.6

2.2.1 Imputation

The dataset presents challenges due to gaps in data caused by intermittent issues with the ICU recording systems. Also, laboratory test results were merged with vital signs which produced a sparse dataset, as tests were done a few times per day. To tackle this, two distinct strategies are employed for imputing missing values, one for vital signs and the other for laboratory values. For vital signs, we initially filled in missing entries by averaging the respective columns for each patient. Then, a moving average (MA) is calculated to minimise the impact of this imputation. When it comes to laboratory tests, linear imputation is used for each patient [10]. An example of this process is demonstrated in Table 1. It is important to note that these imputations are carried out before any further processing of the data.

- For each pair of hourly numerical observations in the dataset, any missing values that occur between them are filled in using linear imputation. This involves assigning values that are equally spaced between the two observations.

$$N = \text{Index}(B_{val}) - \text{Index}(A_{val}) \tag{1}$$

$$X_i = A_{i-1} + \frac{B_{val} - A_{val}}{N} \tag{2}$$

In this context, X_i represents the missing value that needs to be filled in. A_{i-1} is the previous observation. A_{val} is the value of the actual assessment that was recorded before

Table 2 Records are grouped according to the duration of each patient's stay in the ICU

Group No.	Stay duration (h) in ICU
1	<16
2	16 to 30
3	31 to 50
4	51 to 80
5	≥81

Table 3 Each duration of stay was further grouped by age

Group No	Age (Years)
1	15 to 29
2	30 to 39
3	40 to 49
4	50 to 59
5	60 to 69
6	70 to 79
7	80+

the occurrence of X_i, and B_{val} refers to the value of the assessment that was recorded after X_i.

- In cases where a column begins with missing entries, the framework imputes all these initial missing values with the same value as the first subsequent numerical observation in that column.
- When encountering missing values at the end of a column, the procedure involves identifying the last numerical observation. All subsequent missing values are then filled using this last observed numerical value.

Most of the missing values are imputed at this point. However, there could be a case where values for the entire feature are missing for a certain patient. This can happen with laboratory tests. To impute these missing values, the records are categorised based on each patient's duration of stay in the ICU, as detailed in Table 2. Following this, the records within each length-of-stay are further grouped by age, as given in Table 3.

Afterwards, we used mean imputation as described earlier for each age group within each stay duration category instead of individual patients.

2.2.2 Moving average

For vital indicators including heart rate, temperature, diastolic and systolic blood pressure, pulse oximetry (O2Sat), and respiration rate, we computed the moving average (MA) and utilised it for further analysis. However, this approach is not applied to laboratory test values due to their limited frequency of observation—often, there might be only one or two laboratory measurements within a 24-hour period. In technical analysis, a moving average is essentially a kind of limited impulse response filter. It is produced by averaging a number of different subsets of the entire dataset. The literature mentions several types of moving averages, including weighted [24], exponential [5], and simple [24] averages. The conventional form of the moving average of order m was used in our research.

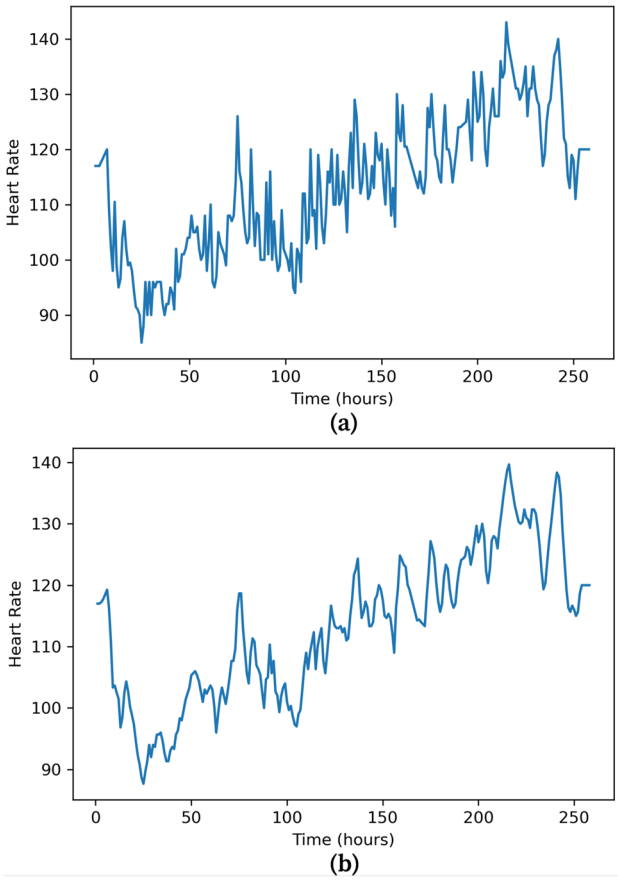

Fig. 3 Following illustrates how the moving average method affects the "Heart Rate" feature for a specific patient: The raw, original data are shown in **a**. The data from **a** are displayed in **b** following the moving average operation

$$\hat{T}_t = \frac{1}{m} \sum_{i=-k}^{k} y_{t+i} \text{ where } m = 2k + 1 \tag{3}$$

The use of moving averages (MAs) helps in eliminating volatility from the data, thereby retaining the trend-cycle component. This is achieved by averaging the values of the time series within k intervals (measured in hours) around a specific time t, which provides an estimation of the trend-cycle at that moment. In our experiments, we tested various values for k, ranging from 1 to 5, to find the most suitable one. We ultimately settled on $k = 2$, as higher values of k tended to overly smooth the data, leading to a flattened curve. Figure 3 demonstrates how the moving average contributes

to smoothing out irregularities or random fluctuations within a feature.

2.2.3 Data normalisation

Data with numerical features are scaled using min–max normalisation. Once the training dataset is normalised, this approach also incorporates the validation and test sets. The significance of this method is becoming more widely acknowledged due to its ability to handle a wide range of data types with distinct features. It addresses numerical problems caused by different ranges of values encountered during computations. The following describes how this scaling method changes a value a into a' within a given range $[\min_{\text{new}}, \max_{\text{new}}]$:

$$a' = \frac{(a - a_{min})}{(a_{max} - a_{min})} \times (\max_{\text{new}} - \min_{\text{new}}) + \min_{\text{new}} \tag{4}$$

where a_{min} and a_{max} are the minimum and maximum values of the original data, respectively. $\max_{\text{new}}$ and $\min_{\text{new}}$ are the desired maximum and minimum values for the new normalised range. In our case, $\min_{\text{new}} = 0$ and $\max_{\text{new}} = 1$.

2.2.4 Sequence generation for LSTM

In our study, we utilised a window size of $w = 6$ to generate data sequences for each patient, where each sequence encapsulates six consecutive hours of data. This setup was critical for training our BiLSTM model to predict the onset of sepsis p hours before it occurs. Specifically, the target variable represents the sepsis status at the $w + p$ hour, coded as 0 for absence and 1 for presence. We explored three prediction horizons with p set at 4, 8, and 12 h, aiming to assess the model's forecasting capabilities at varying intervals. This approach allows for a proactive prediction, potentially offering crucial lead time for clinical interventions.

2.3 Convolutional neural network

In the field of deep learning, a pivotal method known as the convolutional neural network (CNN) was pioneered by LeCun in the early 1990 s [35]. This method is notable for having a distinct design that allows it to consistently identify patterns independent of their orientation or position [1]. It resembles a multilayer perceptron (MLP). The CNN is primarily composed of one or more convolutional layers, each of which has a set of weights and a pooling layer that leads to a fully connected layer [27]. The network can efficiently use the local correlations found in the data to extract and distil important characteristics thanks to its structure.

2.3.1 Convolutional layer

The process described involves a convolution operation, where a dot product is computed for each segment of the input data using a specific kernel. After computing this product, a bias term is added to the result. Subsequently, an activation function is applied, producing a transformed value which forms a feature map. This feature map then serves as the input for subsequent layers of the neural networks [14, 42]. To illustrate, let $X_i = [x_1, x_2, x_3, ..., x_N]$ represent a set of input samples, where N is the total number of samples in the set. Following Eq. 5, the convolution operation processes these samples to produce the output values, effectively capturing important features necessary for further layers of analysis.

$$C_i^{l,j} = h\left(b_j + \sum_{m=1}^{M} W_m^j \cdot X_{i+m-1}^j\right) \tag{5}$$

Here, W_m^j indicates the weight corresponding to the jth feature map at the mth filter index, whereas M specifies the kernel, or filter, size. Index l denotes the particular layer in the network; b denotes the bias connected to the jth feature map; and h is the activation function that adds nonlinearity to that layer.

2.3.2 Batch normalisation

Training data are collected incrementally in batches, which are often unevenly distributed. This variation requires frequent adjustments to network parameters, slowing down convergence. Batch normalisation is implemented following a convolutional layer to mitigate these effects. This method involves calculating the mean (μ_D) and variance (σ_D^2) of each batch, normalising the data to zero mean and unit variance. The normalised data ($\hat{x}_l$) is then adjusted with specific weights and biases to better represent complex functions, as detailed in Eqs. 6 to 9 [25]. The batch normalisation makes the process of coordinating updates across the layers of a neural network easier by reparameterising in this fashion.

$$\mu_D = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{6}$$

$$\sigma_D^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_D)^2 \tag{7}$$

$$\hat{x}_l = \frac{x_i - \mu_D}{\sqrt{\mu_D^2 + \epsilon}} \tag{8}$$

$$y_i = \gamma \hat{x}_l + \beta \tag{9}$$

2.3.3 Max-pooling layer

The pooling or sub-sampling layer reduces data dimensionality. We use a one-dimensional maximum pooling layer following batch normalisation and a one-dimensional convolutional layer. This max-pooling layer compresses the feature set by sampling small rectangular blocks and outputting the highest value from each block [14]. We chose max-pooling for its efficiency in selecting the largest value from adjacent inputs [42]. The pooling operation on a feature map is mathematically expressed as:

$$p_i^{l,j} = \max_{r \in R} \{c_{i \times T+r}^{l,j}\} \tag{10}$$

where R is the pooling window size and T is the stride. After pooling, features are converted into a one-dimensional vector for further processing in classification layers [14]. CNNs differ from other deep learning models like feedforward networks by requiring fewer parameters and less preprocessing, making them efficient for deep learning tasks [44].

2.4 Bidirectional long short-term memory (BiLSTM) networks

A wide range of neural network topologies have been built for useful real-world applications since deep learning is the peak of current machine learning technology. The research described in this paper used a potent deep learning methodology to demonstrate its exceptional and persuasive problem-solving skills. Because of its ability to store knowledge for longer periods of time, this specific mechanism is known as long short-term memory.

The capacity of the LSTM to analyse data and recognise important predictive elements makes it unique in the field of deep learning. It expands upon the framework of the recurrent neural network (RNN). The unique design of LSTM was created by experts to address the "vanishing gradient" problem with RNNs [18]. Its architecture includes an input gate, an output gate, a forget gate, and a memory cell [16]. Within the memory block, the forget gate is controlled by a single neural network layer. Equation 11 [58] is used to compute the activation of this gate.

$$f_t = \sigma(W[x_t, h_{t-1}, C_{t-1}] + b_f) \tag{11}$$

where x_t represents the input sequence. The symbol $h_{(t-1)}$ stands for the output from the preceding block. The previous memory of the LSTM block is indicated by $C_{(t-1)}$. The bias vector is denoted by the word b_f. The logistic sigmoid function is represented by σ, and the weights for each input are provided by W. The input gate serves as a part of a basic neural network that combines the impact of the preceding memory block with the tanh activation function to create new memory. The calculations for these processes are specified in Eqs. 12 and 13 [58].

$$i_t = \sigma(W[x_t, h_{t-1}, C_{t-1}] + b_i) \tag{12}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W[x_t, h_{t-1}, C_{t-1}] + b_c) \tag{13}$$

As long-term dependencies are the default behaviour of LSTM in reality, long-term dependencies may be avoided by deliberately creating and remembering long-term information.

The unidirectional LSTM depends on preceding data, which might not suffice at all times. Conversely, the BiLSTM scrutinises data bi-directionally. The BiLSTM's hidden layer contains a pair of values [19]. One is utilised for computations in the forward phase, and the other for the backward phase. The combined impact of these two variables shapes the BiLSTM's final output, which often improves its anticipated accuracy [12].

2.5 Attention

The attention mechanism is crucial for enhancing model performance by prioritising the most relevant features of the input data [30]. Similar to human attention, which selectively concentrates on certain aspects of our environment while ignoring others, the attention mechanism in this architecture allows the model to differentially weigh various segments of the input. This selective focus facilitates a more nuanced understanding of the complex relationships inherent in the data, which is essential for accurately predicting sepsis at an early stage.

Assuming we have a sequence of numerical input vectors $\mathbf{X} = [x_1, x_2, \ldots, x_N]$, where each x_i is a vector in $\mathbb{R}^d$ (a d-dimensional space), the attention mechanism can be applied as follows:

Generally, we calculate three new vectors for every input vector x_i: a query vector q_i, a key vector k_i, and a value vector v_i. Typically, these are obtained by multiplying x_i by the corresponding weight matrices W_Q, W_K, W_V.

$$q_i = W_Q x_i, \tag{14}$$

$$k_i = W_K x_i, \tag{15}$$

$$v_i = W_V x_i. \tag{16}$$

Here, W_Q, W_K, W_V are the weight matrices to be learned during training.

Subsequently, the attention scores are calculated, which assess how relevant each segment of the input sequence is to the segment that is currently being handled. This is typically

done using a dot product:

$$e_{ij} = q_i^\top k_j, \tag{17}$$

where e_{ij} is the attention score that quantifies the influence of the jth element of the sequence on the ith element.

We utilise a softmax function in order to render these scores manageable and interpretable.

$$\mathbf{s}_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{N} \exp(e_{ik})}, \tag{18}$$

The normalised attention weight in this case is represented by $\mathbf{s}_{ij}$, which denotes the significance of the jth element in the synthesis of the ith element.

Finally, we compute a weighted sum of the value vectors, with the weights being the attention scores:

$$\alpha_i = \sum_{j=1}^{N} s_{ij} v_j, \tag{19}$$

The output vector α_i represents the contextual information gathered throughout the sequence for the ith element by the attention mechanism.

To improve the comprehension and processing of data sequences, this attention mechanism can be incorporated into a variety of models, including LSTM.

2.6 AT-BiLSTM-CNN block

The deep network block leverages the preprocessed data to assess the probability of sepsis occurrence. The architecture of this block, shown in Fig. 4, integrates an attention mechanism, BiLSTM layers, convolutional neural network (CNN) layers, and fully connected (FC) layers. This design facilitates a robust analysis of sequential time series data typical of ICU monitoring systems.

The process begins with each input sequence being directed through an attention layer, which enhances the model's focus on salient features crucial for accurate predictions. The output from this layer is then fed into the BiLSTM layers. These layers are particularly adept at capturing temporal dependencies from both past and future contexts, thanks to their bidirectional processing capability.

Following the BiLSTM layers, the data flow moves to a fully connected (FC) layer, which serves to synthesise the temporal data into a denser representation. This step is critical for integrating the learned temporal features before passing them to the subsequent convolutional layers. These convolutional layers employ a rectified linear unit (ReLU) as an activation function [33], known for its effectiveness in introducing nonlinearity and aiding in faster convergence.

The architecture incorporates two convolutional layers that effectively extract deep characteristics [12, 32] and maintain robustness against noise interference. Employing one-dimensional convolutional layers enables the extraction of location-invariant characteristics from concise time series segments, which is essential for recognising patterns indicative of sepsis. Subsequently, the flattened layer processes the convolutional outputs, transforming them into a format suitable for final classification.

The transformed data is then fed into another fully connected layer which consolidates the extracted features into a final output vector, z. This vector is processed through the following equation to determine the probability of sepsis onset:

$$y = \mathrm{ReLU}(W_y z + b_y) \tag{20}$$

In this equation, W_y and b_y represent the weight matrix and bias, respectively, while y indicates the likelihood of sepsis onset. The model applies an adjustable threshold to decide whether the patient's condition suggests a risk of becoming septic. Training of the network is accomplished using the adaptive moment estimation (ADAM) optimisation algorithm [31], which iteratively updates the network's weights based on the training data to enhance predictive accuracy. The complete training and prediction process is encapsulated in Algorithm 1, ensuring a systematic approach to model training and deployment.

This architecture, combining attention mechanisms with BiLSTM and CNN layers, is tailored to provide early and reliable predictions of sepsis, potentially allowing for timely clinical interventions.

3 Empirical evaluation

3.1 Data analysis

An analysis of the dataset unveiled the count of male and female participants, as well as the prevalence of sepsis and non-sepsis conditions within each gender group. The study cohort consisted of 55.9% males and 44.1% females. Furthermore, when collecting data, sepsis was detected in 7.7% of the male individuals and 6.7% of the female participants. The processed dataset consists of about 1,542,641 h of monitored data gathered from ICUs.

Table 4 presents the statistical summaries—including the mean $\pm$ std and median values of vital signs and laboratory test results within the dataset, offering insights into the distribution of these measures. The dataset encompasses demographic details such as age, sex, duration between hospital admission and transfer to the ICU, the type of ICU, and the duration of the ICU stay. Figure 5a delineates the duration of ICU stays in hours, with an average stay amounting

to 51.99 ± 42.75 h, and identifies a sepsis occurrence rate of 7.2%.

Algorithm 1 Train attention-based BiLSTM-CNN model for sepsis prediction

```
Inputs:
    data: Time-series clinical data for patients
    labels: Disease labels corresponding to each sequence in the data
    epochs: Number of training epochs
    batch_size: Size of each batch for training
Outputs:
    A trained AT-BiLSTM-CNN model
    Performance metrics on the test data set
procedure TRAIN AT-BILSTM-CNN MODEL(data, labels, epochs, batch_size)
    Initialize the AT-BiLSTM-CNN model with desired architecture
      Attention layer for capturing relationships within the data
      BiLSTM layers for capturing time dependencies
      FC layers to process the outputs of BiLSTM layers
      CNN layers for feature extraction
      FC layers for prediction
    Preprocess the data
      Normalize the clinical features
      Split the data into training, validation, and test sets
      Convert data into sequences suitable for BiLSTM input
    for epoch in 1 to epochs do
      for each batch in training data do
          Prepare input sequences and corresponding labels for the batch
          Forward pass: Feed the input sequences through the model
          Compute loss: Compare the prediction with the true labels
          Backward pass: Calculate gradients with respect to the loss
          Update model weights: Apply ADAM optimizer to update weights
      end for
      Evaluate model on validation set
        Compute validation loss and accuracy
        If validation accuracy improves, save the model as the best model
    end for
    Load the best saved model
    Test the model on the test set
      Compute test loss and accuracy
      Report test metrics for model performance
end procedure
```

Furthermore, Fig. 5b depicts the age distribution and the incidence of sepsis across different age groups, highlighting that 58.6% of patients admitted to the ICU are aged above 60.

3.2 Performance analysis

We randomly partitioned our dataset into a training set, encompassing 80% of the patient data, and a test set, consisting of the remaining 20%. It is crucial to note that each record contains hourly data for a single patient. To ensure the reliability of our evaluation, we utilised a stratified tenfold cross-validation technique. This approach divides the data into ten segments of identical size, guaranteeing that each segment has a comparable distribution of sepsis cases. During each validation cycle, one segment is allocated for testing purposes, while the remaining nine segments are utilised for training. The aforementioned cycle is repeated ten times, wherein each section is utilised once as the test set. We calculate the mean of the outcomes to derive a final assessment. To rectify disparities in data classes, we also incorporated window slicing augmentation [57], which involves the selection of random, contiguous segments from the electronic health records of patients.

Two techniques in our deep learning model counteract overfitting: dropout, which randomly removes units during training, and l^2 weight decay, which limits the l^2 norm of the weights. The ADAM optimiser is employed to update the network weights, with the process being carried out by shuffling through batches of 64 samples. In order to enhance the mean squared error (MSE) loss function, we utilise the Nesterov acceleration technique [48]. The duration of our training lasted for 1000 epochs.

Python with PyTorch framework is utilised for building the model. The computations were performed using a GeForce RTX 3080 NVIDIA GPU powered by an i7 processor with 64 GB RAM.

Figure 6 demonstrates the performance of AT-BiLSTM-CNN compared to the Sepsis-3 criteria [51] for a specific patient. The patient promptly meets the Sepsis-3 criteria. However, the AT-BiLSTM-CNN model exceeded the crucial threshold almost twelve hours before the occurrence of sepsis. This showcases the model's capacity to accurately identify intricate patterns in clinical features and predict an imminent risk of sepsis in advance.

We assessed the accuracy, sensitivity, specificity, and AUROC (area under the receiver operating characteristic) of our model [55]. In order to establish similarities, we replaced the BiLSTM component in our model's structure with a gated recurrent unit (GRU) [9]. In addition, we constructed a random forest (RF) classifier that utilises a predictive goal. This classifier was trained using data from a certain time point, t, to forecast outcomes at $t + 4$, $t + 8$, and $t + 12$ hours. The results for the predicted periods of 4, 8, and 12 h are summarised in Table 5. The findings demonstrate that BiLSTM surpasses both GRU and RF in terms of performance. This test assesses the proficiency of the AT-BiLSTM-CNN model.

During our second experiment, we conducted a comparison of the results obtained from InSight [13] and AISE [41], which are two state-of-the-art methodologies. Significantly, the accuracy of our model decreased as the prediction window increased. Table 6 illustrates the performance of AT-BiLSTM-CNN, AISE, and InSight under different forecast intervals. It is evident that AT-BiLSTM-CNN performs exceptionally well, especially when predicting outcomes within a 4-hour timeframe.

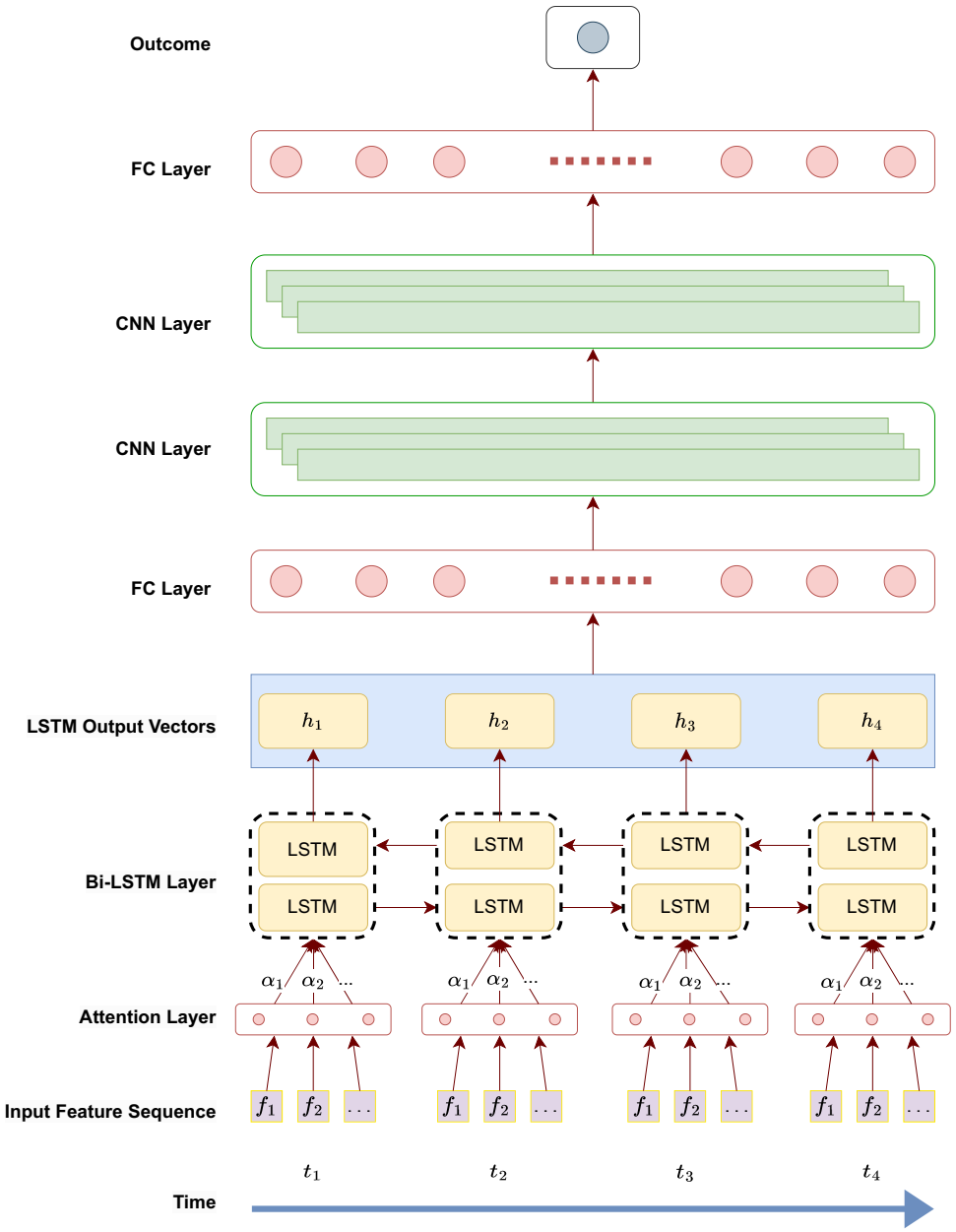

Fig. 4 Architecture of proposed AT-BiLSTM-CNN model

Subsequently, we assessed our proposed model by excluding its one-dimensional convolutional layers in order to measure the influence of our architectural design. The results, displayed in Table 7, indicate a significant decrease in performance relative to the complete model.

3.3 Discussion

The aim of our study is to present a AT-BiLSTM-CNN architecture that may be used to detect sepsis at an early stage. We employed the ICU patient records from the 2019 PhysioNet/Computing in Cardiology Challenge for the purposes of training, validation, and testing.

A detailed comparison of three predictive models-AT-BiLSTM-CNN, AT-GRU-CNN, and RF is shown in Table 5. AUROC shows the ability of the model to discriminate between classes. The AT-BiLSTM-CNN tends to perform best at shorter time frames, with performance decreasing at twelve hours. The AT-GRU-CNN and RF show similar trends, although the RF model's performance drops more noticeably at twelve hours. The AT-BiLSTM-CNN maintains relatively stable sensitivity across time frames, while the AT-GRU-CNN and RF exhibit a decline as the prediction window increases. All three models maintain relatively stable specificity, with minor fluctuations. This suggests these models are consistent in correctly identifying negatives across differ-

Table 4 Statistical summary of clinical features which are used in our work

Feature	Unit	Mean ± STD			Median		
		Sepsis	Non-sepsis	Total	Sepsis	Non-sepsis	Total
Heart rate	Beats/min	90.9 ± 19.1	84.3 ± 17.3	84.4 ± 17.4	90.0	83.6	83.7
Temperature	Deg C	37.1 ± 1.0	36.9 ± 0.7	36.9 ± 0.7	37.1	36.8	36.8
O_2 Saturation	%	97.0 ± 3.4	97.2 ± 3.1	97.2 ± 3.1	98.0	97.6	97.6
Diastolic pressure	mm Hg	62.4 ± 14.1	64.2 ± 14.1	64.1 ± 14.1	61.0	63.3	63.3
Systolic pressure	mm Hg	121.5 ± 25.0	123.8 ± 23.2	123.7 ± 23.3	118.5	122.4	122.4
Respiration	Breaths/min	20.2 ± 6.3	18.6 ± 5.1	18.7 ± 5.1	20.0	18.3	18.3
Glucose	mg/dL	136.4 ± 49.2	131.2 ± 45.7	131.3 ± 45.8	126.0	126.6	126.6
Leukocyte	count*$10^3/\mu L$	13.0 ± 8.8	11.1 ± 6.7	11.1 ± 6.8	11.7	10.6	10.7
Creatinine	mg/dL	1.7 ± 1.8	1.4 ± 1.7	1.4 ± 1.7	1.1	1.4	1.4
Platelet	count*$10^3/\mu L$	204.6 ± 119.8	204.8 ± 101.8	204.8 ± 102.1	183.0	196.9	196.9
Haemoglobin	g/dL	10.2 ± 1.9	10.6 ± 2.0	10.6 ± 2.0	10.0	10.5	10.5
Haematocrit	%	30.7 ± 5.4	31.7 ± 5.6	31.7 ± 5.6	30.1	31.5	31.5
BUN	mg/dL	29.5 ± 22.6	22.8 ± 18.7	22.9 ± 18.8	22.4	20.7	20.8

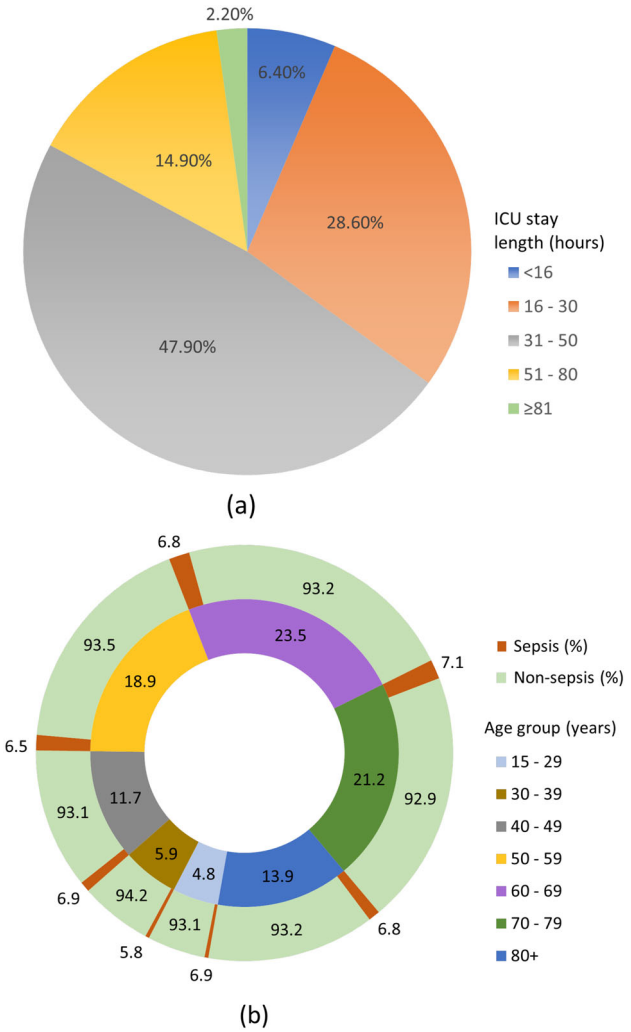

Fig. 5 Distribution of **a** ICU stay duration and **b** patient's age along with the incidence of sepsis across different age groups

ent time frames. AT-BiLSTM-CNN and AT-GRU-CNN show good accuracy at four hours, with a slight decrease at twelve hours. RF has the lowest accuracy at twelve hours, which corresponds with its lower AUROC, sensitivity, and specificity scores. Overall, the AT-BiLSTM-CNN model exhibits robust performance across all metrics compared to the AT-GRU-CNN and RF models, particularly in longer prediction windows.

We compared our model to the AISE [41] and InSight [13] approaches, as shown in Table 6. AISE demonstrated AUROCs of 0.85, 0.84 and 0.83 for sepsis predictions made four, eight and twelve hours in advance, respectively. In comparison, InSight produced an AUROC of 0.74 when predicting sepsis four hours ahead.

In literature, Scherpf et al. [49] achieved an AUROC of 0.81 by utilising a recurrent neural network (RNN) with a three-hour advantage. Lauritsen et al. [34] used the SIRS criteria and found AUROCs of 0.856 and 0.756 for predicting sepsis at three and twenty-four hours before its onset, respectively. Our AT-BiLSTM-CNN model effectively utilises the modern Sepsis-3 criteria, moving beyond older benchmarks like SIRS [51].

The AT-BiLSTM-CNN model exhibited a significant level of accuracy in its predictions, with an average AUROC of 0.88 for four-hour forecasts and maintaining its resilience with an average AUROC of 0.84 for twelve-hour forecasts. The comprehensive data and comparative studies presented in Table 6 provide evidence of the improved predictive capabilities of AT-BiLSTM-CNN.

From Table 7, it is evident that there is a consistent decline in all performance metrics as the prediction time frame extends, which is typical in predictive modelling as longer-term predictions are generally more challenging and

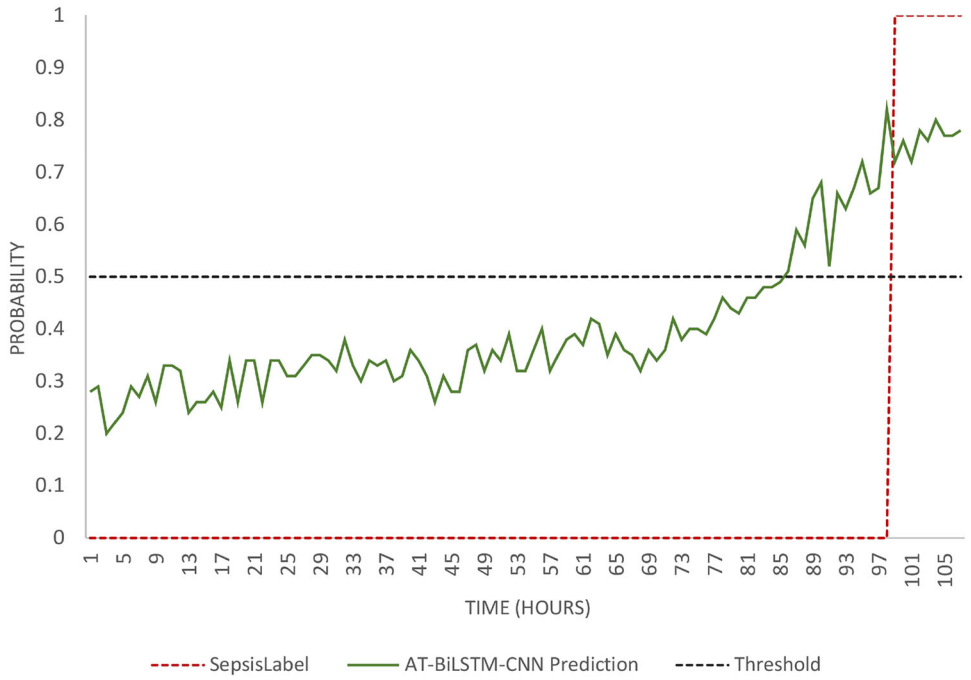

Fig. 6 Comparison of the AT-BiLSTM-CNN prediction with Sepsis-3

Table 5 Performance metrics of AT-BiLSTM-CNN and GRU-CNN and RF model at different time frames

	AT-BiLSTM-CNN			AT-GRU-CNN			RF		
Performance metric	4 h	8 h	12 h	4 h	8 h	12 h	4 h	8 h	12 h
AUROC	0.88±0.044	0.85±0.026	0.84±0.029	0.85±0.026	0.82±0.037	0.79±0.015	0.82±0.018	0.78±0.036	0.74±0.041
Sensitivity	0.84±0.031	0.81±0.016	0.80±0.022	0.80±0.026	0.78±0.045	0.74±0.039	0.74±0.045	0.69±0.058	0.63±0.049
Specificity	0.89±0.082	0.87±0.038	0.86±0.040	0.89±0.028	0.87±0.047	0.84±0.026	0.89±0.016	0.87±0.030	0.85±0.039
Accuracy	0.90±0.076	0.87±0.037	0.86±0.040	0.89±0.028	0.87±0.046	0.84±0.025	0.88±0.015	0.86±0.028	0.84±0.034

Table 6 Comparative results of AT-BiLSTM-CNN, AISE, and InSight at different time frames

	AT-BiLSTM-CNN			AISE			InSight		
Performance metric	4 h	8 h	12 h	4 h	8 h	12 h	4 h	8 h	12 h
AUROC	0.88	0.85	0.84	0.85	0.84	0.83	0.74	–	–
Sensitivity	0.84	0.81	0.80	–	–	–	0.80	–	–
Specificity	0.89	0.87	0.86	0.67	0.65	0.63	0.54	–	–
Accuracy	0.90	0.87	0.86	0.67	0.66	0.63	0.57	–	–

prone to uncertainty. The findings indicate that the exclusion of CNN layers from this architecture detrimentally affects performance across all time frames. The negative impact becomes more pronounced specially for AUROC and sensitivity as the duration of the prediction window increases.

The AT-BiLSTM-CNN is particularly effective in handling detailed time series data, making it well-suited for applications that include complex EHRs. It is very suitable for developing expert systems that handle such records.

3.4 Limitations and future work

The interaction between doctors and advanced systems such as deep learning models highlights the importance of using models that can be easily understood in clinical settings. Shickel et al. [50] emphasised that neglecting the significance of these models in assisting physicians could be a crucial mistake. This risk is heightened due to the fact that models such as AT-BiLSTM-CNN are not designed to detect specific signs and rules for manual scoring. Instead, they make predictions using intricate computations in a multi-dimensional feature space.

Table 7 Performance metrics produced by our architecture without CNN layers at different time frames

Performance metric	4 h	8 h	12 h
AUROC	0.85	0.83	0.81
Sensitivity	0.81	0.78	0.76
Specificity	0.88	0.86	0.85
Accuracy	0.88	0.86	0.85

When using deep learning models in medical settings, it is crucial to recognise and address any biases due to the complicated nature of their operation. Research suggests that these models may not have the same level of applicability as previously assumed, due to differences in processes and the individual hardware and software utilised [2]. Therefore, the implementation of the AT-BiLSTM-CNN in unfamiliar settings requires careful and thoughtful preparation.

To mitigate some of these concerns, the incorporation of methods like SHAP (SHapley Additive exPlanations) [36] or LIME (Local Interpretable Model-agnostic Explanations) [47] can be instrumental. These techniques provide insights into the feature importance and offer explainable predictions, which can help clinicians understand how the models make their decisions. By clarifying which features are most influential in predicting sepsis, healthcare professionals can gain a better understanding of the model's reasoning, potentially increasing trust and facilitating more effective interactions between the model and clinicians.

The dataset's limitation of excluding patients who acquire sepsis within four hours of ICU admission also restricts the application of our technique to those who are already septic upon entering the ICU. Furthermore, due to the absence of data about sepsis-related comorbidities, our approach does not include unorganised data such as medical histories, comorbidities, or doctor's notes. These limitations could be the subject of future research.

4 Conclusion

Anticipating sepsis at an early stage continues to be an intricate obstacle, frequently impeding clear identification until it progresses significantly. We have devised a sophisticated machine learning framework called AT-BiLSTM-CNN, which is specifically designed to forecast the onset of sepsis with a lead time of up to 12 h. This model demonstrates robustness against missing data, extreme values, and is capable of providing accurate predictions even in the presence of random fluctuations and inaccuracies. We conducted training, testing, and validation of the model using data that adhered to the Sepsis-3 criteria.

Sepsis is particularly fatal in intensive care unit (ICU) patients when it remains undetected or is incorrectly identified. Deploying such sophisticated instruments is imperative in clinical environments, as it has the potential to diminish the economic expenses, intensive care unit durations, and the significant morbidity and mortality linked to sepsis. The encouraging results indicate that the AT-BiLSTM-CNN model has the potential to greatly aid healthcare workers in predicting sepsis in ICU environments.

Acknowledgements We gratefully acknowledge the support of the ELISE project in the completion of this work. This project was partially funded by the Federal Ministry of Health under Grant No. 2520DAT66A.

References

1. Acharya, U.R., Shu Lih, O., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., San, T.R.: A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. **89**, 389–396 (2017)
2. Agniel, D., Kohane, I.S., Weber, Griffin M.: Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ, 361 (2018)
3. Alam, N., Hobbelink, E.L., van Tienhoven, A.-J., van de Ven, P.M., Jansma, E.P., Nanayakkara, P.W.B.: The impact of the use of the early warning score (ews) on patient outcomes: a systematic review. Resuscitation **85**(5), 587–594 (2014)
4. Bone, R.C., Balk, R.A., Cerra, F.B., Phillip Dellinger, R., Fein, A.M., Knaus, W.A., Schein, R.M.H., Sibbald, W.J.: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Chest **101**(6), 1644–1655 (1992)
5. Brown, R.G.: Exponential smoothing for predicting demand. Operat. Res. **5**, 145–145 (1957)
6. Calvo, M., Jané, R.: Sleep stage influence on the autonomic modulation of sleep apnea syndrome. In: 2019 Computing in Cardiology (CinC), 1–4. IEEE (2019)
7. cdc.gov. https://www.cdc.gov/sepsis/datareports/index.html. [Accessed 14-11-2023]
8. Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y.: Recurrent neural networks for multivariate time series with missing values. Sci. Rep. **8**(1), 6085 (2018)
9. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling arXiv preprint (2014) arXiv:1412.3555
10. Das, P.P., Mast, M., Wiese, L., Jack, T., Wulf, A.: Data extraction for associative classification using mined rules in pediatric intensive care data. BTW 2023 (2023)
11. Davoudi, A., Malhotra, K.R., Shickel, B., Siegel, S., Williams, S., Ruppert, M., Bihorac, E., Ozrazgat-Baslanti, T., Tighe, P.J., Bihorac, A., et al.: Intelligent ICU for autonomous patient monitoring using pervasive sensing and deep learning. Sci. Rep. **9**(1), 8020 (2019)
12. De Baets, L., Ruyssinck, J., Peiffer, T., Decruyenaere, J., De Turck, F., Ongenae, F., Dhaene, T.: Positive blood culture detection in time series data using a bilstm network (2016) arXiv preprint arXiv:1612.00962
13. Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., Shimabukuro, D., Chettipally, U., Feldman, M.D., Barton, C., et al.: Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med. Inform. **4**(3), e5909 (2016)

14. Dmitrievich, I. A.: Deep learning in information analysis of electrocardiogram signals for disease diagnostics. The Ministry of Education and Science of The Russian Federation Moscow Institute of Physics and Technology (2015)
15. Dorsett, M., Kroll, M., Smith, C.S., Asaro, P., Liang, S.Y., Moy, H.P.: Gsofa has poor sensitivity for prehospital identification of severe sepsis and septic shock. Prehosp. Emerg. Care **21**(4), 489–497 (2017)
16. El-Shafiey, M.G., Hagag, A., El-Dahshan, E.-S.A., Ismail, M.A.: A hybrid bidirectional LSTM and 1d CNN for heart disease prediction. IJCSNS **21**(10), 135 (2021)
17. Faisal, M., Scally, A., Richardson, D., Beatson, K., Howes, R., Speed, K., Mohammed, M.A.: Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. Crit. Care Med. **46**(4), 612–618 (2018)
18. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: Continual prediction with lstm. Neural Comput. **12**(10), 2451–2471 (2000)
19. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)
20. Halpern, N.A., Pastores, S.M.: Critical care medicine in the united states 2000–2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. Crit. Care Med. **38**(1), 65–71 (2010)
21. Herasevich, V., Kor, D.J., Subramanian, A., Pickering, B.W.: Connecting the dots: rule-based decision support systems in the modern emr era. J. Clin. Monit. Comput. **27**, 443–448 (2013)
22. Horng, S., Sontag, D.A., Halpern, Y., Jernite, Y., Shapiro, N.I., Nathanson, L.A.: Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS ONE **12**(4), e0174708 (2017)
23. Hye, J.K., Ha, Y.K.: Learning representations for the early detection of sepsis with deep neural networks. Comput. Biol. Med. **89**, 248–255 (2017)
24. Hyndman, R.J.: Moving averages, pp. 866–869. Springer, Berlin and Heidelberg (2011)
25. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, 448–456. pmlr (2015)
26. Jalali, A., Bender, D., Rehman, M., Nadkanri, V., Nataraj, C.: Advanced analytics for outcome prediction in intensive care units. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 2520–2524 (2016)
27. Johnson, J., Karpathy, A., Li, F.-F.: CS231n convolutional neural networks for visual recognition. Notes of Stanford CS class (2016). https://cs231n.stanford.edu/2016/
28. Kamaleswaran, R., Akbilgic, O., Hallman, M.A., West, A.N., Davis, R.L., Shah, S.H.: Applying artificial intelligence to identify physiomarkers predicting severe sepsis in the picu. Pediatr. Crit. Care Med. **19**(10), e495–e503 (2018)
29. Kashiouris, M., O'Horo, J.C., Pickering, B.W., Herasevich, V.: Diagnostic performance of electronic syndromic surveillance systems in acute care. Appl. Clin. Inform. **4**(02), 212–224 (2013)
30. Kim, Y., Denton, C., Hoang, L., Rush, A.M.: Structured attention networks (2017) arXiv preprint arXiv:1702.00887
31. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2014) arXiv preprint arXiv:1412.6980
32. Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J.: 1d convolutional neural networks and applications: a survey. Mech. Syst. Signal Process. **151**, 107398 (2021)
33. Koehler, F., Risteski, A.: Representational power of relu networks and polynomial kernels: beyond worst-case analysis (2018) arXiv preprint arXiv:1805.11405
34. Lauritsen, S.M., Kalør, M.E., Kongsgaard, E.L., Lauritsen, K.M., Jørgensen, M.J., Lange, J., Thiesson, B.: Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif. Intell. Med. **104**, 101820 (2020)
35. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Comput. **1**(4), 541–551 (1989)
36. Lundberg, S., Lee, S.-I.: A unified approach to interpreting model predictions (2017)
37. Mao, Q., Jay, M., Hoffman, J.L., Calvert, J., Barton, C., Shimabukuro, D., Shieh, L., Chettipally, U., Fletcher, G., Kerem, Y., et al.: Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and icu. BMJ Open **8**(1), e017833 (2018)
38. Martin, G.S.: Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. Expert Rev. Anti Infect. Ther. **10**(6), 701–706 (2012)
39. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T.: Deep learning for healthcare: review, opportunities and challenges. Brief. Bioinform. **19**(6), 1236–1246 (2018)
40. Mollura, M., Mantoan, G., Romano, S., Lehman, L.-W., Mark, R.G., Barbieri, R.: The role of waveform monitoring in sepsis identification within the first hour of intensive care unit stay. In: 2020 11th conference of the European study group on cardiovascular oscillations (ESGCO). IEEE, pp. 1–2(2020)
41. Nemati, S., Holder, A., Razmi, F., Stanley, M.D., Clifford, G.D., Buchman, T.G.: An interpretable machine learning model for accurate prediction of sepsis in the icu. Crit. Care Med. **46**(4), 547 (2018)
42. Nurmaini, S., Umi Partan, R., Caesarendra, W., Dewi, T., Naufal Rahmatullah, M., Darmawahyuni, A., Bhayyu, V., Firdaus, F.: An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique. Appl. Sci. **9**(14), 2921 (2019)
43. Paoli, C.J., Reynolds, M.A., Sinha, M., Gitlin, M., Crouser, E.: Epidemiology and costs of sepsis in the united states-an analysis based on timing of diagnosis and severity level. Crit. Care Med. **46**(12), 1889 (2018)
44. Pourbabaee, B., Roshtkhari, M.J., Khorasani, K.: Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. IEEE Trans. Syst., Man, Cybern.: Syst. **48**(12), 2095–2104 (2017)
45. Reyna, Matthew A., Josef, C., Seyedi, S., Jeter, R., Shashikumar, S.P., Westover, M.B., Sharma, A., Nemati, S., Clifford, G.D.: Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. In 2019 Computing in Cardiology (CinC), 1. IEEE (2019)
46. Reyna, M.A., Josef, C.S., Jeter, R., Shashikumar, S.P., Brandon Westover, M., Nemati, S., Clifford, G.D., Sharma, A.: Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. Crit. Care Med. **48**(2), 210–217 (2020)
47. Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should i trust you?": explaining the predictions of any classifier (2016)
48. Ruder, S.: An overview of gradient descent optimization algorithms (2016) arXiv preprint arXiv:1609.04747
49. Scherpf, M., Gräßer, F., Malberg, H., Zaunseder, S.: Predicting sepsis with a recurrent neural network using the mimic iii database. Comput. Biol. Med. **113**, 103395 (2019)
50. Shickel, B., Tighe, P.J., Bihorac, A., Rashidi, P.: Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. IEEE J. Biomed. Health Inform. **22**(5), 1589–1604 (2017)
51. Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G.R., Chiche, J.-D., Coopersmith, C.M., et al.: The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA **315**(8), 801–810 (2016)

52. Smith, G.B., Prytherch, D.R., Meredith, P., Schmidt, P.E., Featherstone, P.I.: The ability of the national early warning score (news) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation **84**(4), 465–470 (2013)
53. Subbe, C.P., Kruger, M., Rutherford, P., Gemmel, L.: Validation of a modified early warning score in medical admissions. QJM **94**(10), 521–526 (2001)
54. Torio, C.M., Moore, B.J.: National inpatient hospital costs: the most expensive conditions by payer, 2013: statistical brief# 204. Healthcare cost and utilization project (HCUP) statistical briefs, 2006–2016 (2006)
55. Trevethan, R.: Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front. Public Health **5**, 307 (2017)
56. Usman, O.A., Usman, A.A., Ward, M.A.: Comparison of sirs, qsofa, and news for the early identification of sepsis in the emergency department. Am. J. Emerg. Med. **37**(8), 1490–1497 (2019)
57. Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., Xu, H.: Time series data augmentation for deep learning: a survey (2020) arXiv preprint arXiv:2002.12478
58. Yildirim, Ö.: A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput. Biol. Med. **96**, 189–202 (2018)