

Multimodal Deep Learning Framework for Early Sepsis Prediction Using ICU Physiological and Clinical Data

Kusuma S*, Karthik Reddy M, Kiran Kumar M

Department of Computer Science and Engineering (Data Science), Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh, India. *Corresponding Author's Email: kusu87@gmail.com

Abstract

Sepsis is a major cause of mortality worldwide, particularly in Intensive Care Units (ICUs), where timely diagnosis and early prediction are essential for improving patient outcomes. Sepsis is a major cause of death in the world and more so in intensive care units (ICUs) where a timely given diagnosis and prompt prognostication are imminent factors in death. Traditional clinical scoring procedures and many machine learning procedures often do not include physiological biomarkers with dynamic temporal changes that determine the development of diseases, which results in limiting their predictive value. To counter this, the current study presents a multitask deep learning network aimed at predicting the development of sepsis based on longitudinal physiological and clinical data that are collected among ICU patients. The suggested architecture 1 combines heterogeneous electronic health record components: vital sign traces, laboratory analysis results and demographic factors into consistently divided time series to forward them to high-level sequence-sensitive deep learning models namely Temporal Convolutional Network (TCN) and a coupled Convolutional Neural Network Long Short-Term Memory (CNN-LSTM) architecture. Temporal masking and median-imputed methods are used in order to attenuate the impacts of the skipped entries and varying sequence lengths. Evaluation of the predictive accuracy visually empirically demonstrates that the TCN has reached the highest predictive accuracy of 0.9785, followed by CNN-LSTM with 0.9730 and a gradient boosting ensemble that has reached 0.9690. These results emphasize the substantive benefit that exists as a result of temporal modeling in the improvement of the predictive systems about early sepsis detection.

Keywords: Clinical Time-series Analysis, CNN-LSTM Models, Deep Learning, ICU Analytics, Sepsis Prediction, Temporal Convolutional Networks.

Introduction

Sepsis is a serious condition characterized by the body's excessive reaction to infections, resulting in organ dysfunction (1). It is currently one of the biggest causes of death and morbidity, especially in critically ill people admitted to Intensive Care Units (ICUs) (2). Even though many advancements have been made in antimicrobial treatment, critical care medicine and hemodynamic monitoring, sepsis continues to pose a huge global challenge due to its rapid onset, heterogeneity and multi-organ failure (3, 4). Late detection has been shown to result in higher mortality, long hospital stays and higher healthcare costs (5). Therefore, early and accurate prediction of sepsis onset before any irreparable physiological damage has become one of the biggest challenges in the field of critical care medicine (6). Current diagnostic approaches to sepsis depend primarily on the use of clinical score systems, including the SOFA, quick SOFA (qSOFA) and Modified Early Warning Score (MEWS) (1).

Though the aforementioned methods provide objective criteria for the diagnosis of the condition, they are mainly based on fixed thresholds and static physiological measures (1). As a result, they do not take into account any changes in patient physiological status prior to the development of clinical signs of deterioration, which limits their capacity to provide early alerts. Moreover, due to variability of disease presentations in sepsis, rule-based score systems lack both sensitivity and specificity, thus requiring the development of a more adaptive data-driven solution (7). The recent development of healthcare in terms of its digital transformation provided researchers with a huge amount of information collected in the form of electronic health records (8). Such data include vital signs, lab results, medication history, demographics and clinical observations (9). All those heterogeneous sources of data allow the creation of AI-based models for continuous

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution and reproduction in any medium, provided the original work is properly cited.

(Received 12th February 2026; Accepted 25th August 2026; Published 11th October 2026)

monitoring of patients' state (10). ML algorithms like logistic regression, SVM, Random Forest and Gradient Boosting have been used successfully in order to learn complex nonlinear relations between clinical features (11, 12). Nevertheless, the majority of traditional methods consider patient observations independently without taking into account temporal dynamics of physiological variables, which is a crucial factor for sepsis prognosis (12).

Deep learning technology is recognized as a revolutionary technique in healthcare analytics due to the ability to learn hierarchical feature representation directly from high-dimensional clinical data (13). Various architectures like RNNs, LSTM networks, GRUs and CNNs were studied for modeling sequential physiological processes related to the onset of sepsis (14, 15). Superior results obtained through these models compared to classical machine learning methods are due to the ability to capture complex interactions and dependencies among many clinical variables (16). Nonetheless, recurrent architectures face challenges like exploding and vanishing gradients, high computational requirements and inefficiency of learning from long clinical time series (17). Besides, ICU datasets are characterized by highly uneven distribution of observations, irregularity of the sampling period, varying lengths of sequences and class imbalance (18).

Temporal deep learning models recently became an area of great interest for addressing temporal dependencies among clinical variables. Using causality and dilation mechanisms allows TCNs to learn long-range temporal dependencies while being computationally efficient because of their ability to perform parallel computations and stable gradient propagation (19). The difference from the traditional recurrent models is that TCNs can work with the whole sequence at once, which provides better computational efficiency and scalability (20). At the same time, multimodal learning techniques using information from physiological signals, laboratory analysis and demographic features show better results than unimodal approaches (21, 22). Further developments in explainable artificial intelligence, federated learning, transfer learning and hybrid deep learning architectures demonstrate the potential of AI-based clinical decision support systems in predicting sepsis (23). However, there are still

challenges related to the interpretation of models, external validation, clinical implementation and processing of irregular time series.

Therefore, one of the most important research gaps in this field is the creation of a predictive framework that can address the issue of modeling long-range temporal dependencies, multimodal learning and processing of missing observations and variable-length sequences, as well as provide computational efficiency required for real-time clinical applications (22). Current studies focus on prediction performance while not considering much the issues related to the sequence-aware learning approach and preservation of the continuity of sequences during training (19). Besides, using padding during construction of sequences of fixed length might cause bias if invalid data are used in the optimization process (18). It is necessary to develop architectures that could learn robust temporal representations and distinguish valid data from the padded data. In order to solve the above problems, in this study, a sequence-aware multimodal deep learning approach is proposed for early sepsis prediction from ICU clinical time-series data. Specifically, the TCNs will be combined with a hybrid architecture that incorporates both convolutional neural networks and LSTM in order to discover complex temporal relationships between multiple heterogeneous physiological, laboratory and demographic attributes.

This study will have four main goals:

- (i) to design a sequence-aware, multimodal deep learning architecture for early sepsis prediction using ICU clinical time-series data;
- (ii) to comprehensively compare the predictive performance of proposed (TCN), traditional (CNN-LSTM) and gradient boosting models;
- (iii) to study the effects of temporal sequence modeling and temporal masking on prediction performance and robustness; and
- (iv) to evaluate the practical feasibility of applying sequence-aware deep learning models for clinical decision support systems in ICU settings.

Methodology

The proposed methodology includes a sequence modeling framework to learn from ICU time-series clinical data using sequence modeling deep learning algorithms while addressing the

challenges of variable-length sequences, missing data and padding bias (19). The proposed methodology consists of constructing temporal sequences, designing a model architecture, creating a masked loss function and determining the model training strategy. Two different types of sequence-based neural network models will be

trained, evaluated and compared with a third, gradient boosting classifier. Temporal masking is a key element of the methodology to avoid influence from padding added to time steps during their construction during the process of establishing an unbiased and fair learning process for clinical data time series that can vary in length (18).

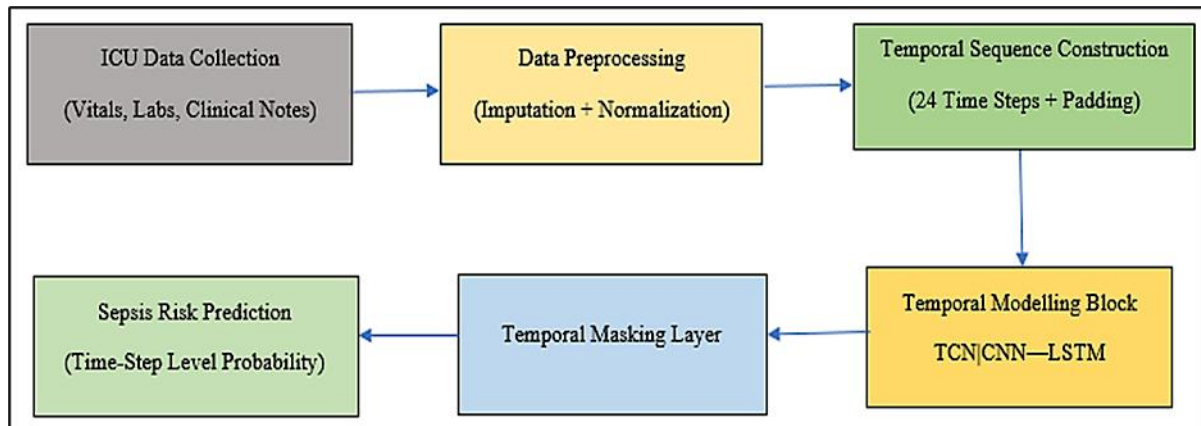


Figure 1: The Proposed Multimodal Deep-learning Framework was Developed to Predict Sepsis at the Early Stage; System Architecture is Outlined

Figure 1 presents an architecture of a multimodal deep learning system designed to predict sepsis from physiological and clinical ICU data. The flow starts at the collection of data (vital signs, lab data and clinical notes) and undergoes preprocessing that consists of median imputation and normalization. Furthermore, the data are reconstructed to be of a fixed length and varying length patient data is dealt with by a technique referred to as temporal masking. These fixed length data sets are then subjected to extraction of temporal features via TCN and CNN - LSTM. The developed framework provides a stepwise prediction of sepsis risk, which supports the provision of timely and critical clinical decisions within the context of sequence aware deep learning.

Data Source and Collection

Research on the proposed method is based on clinical data collected from intensive care units (ICUs) and available through electronic health records (EHRs) (8). The data includes multiple variables representing physiological and clinical measurements taken during an ICU patient's stay in the hospital. Each entry in the dataset reflects changes in a patient's physiology at a unique point in time. The dataset contains regularly measured vital signs, lab tests and clinical factors frequently used to evaluate patients with sepsis. The dataset

also includes a binary variable to indicate whether or not sepsis is present (per standard clinical definition) (1). All patient identifiers have been anonymized for privacy purposes and in accordance with ethical data use guidelines.

Dataset Description

Clinical ICU Time Series Datasets for Experimental Evaluation Time-series datasets containing clinical information have been generated from electronic health records collected from ICUs. The dataset included multiple types of physiological measurement variables, laboratory tests and demographic information for each patient who was treated in an ICU. The clinical time-series datasets have been organized into 24 time steps (fixed-length sequences) for each patient to support temporal modeling. Each time step represents data collected at a specific time point and the presence or absence of sepsis (e.g., positive or negative) is indicated by a binary label. When restructuring patient records into fixed-length sequences, some patients had fewer than 24 time points, so additional zeros were added as placeholders until a total of 24 time points were achieved; these additional time points were then color-coded (using a temporal mask) to show which time points were valid observations and which were zeros (placeholders).

Clinical datasets usually contain a high volume of missing or infrequently sampled values due to the variability of measurement frequency, resulting in missing or infrequent data (18). Invalid numerical values were replaced with missing (N/A) and median imputation was used to fill those in. The imputation method chosen was based on its robustness to outliers and preservation of the underlying distribution of clinical variables. After imputation, each clinical variable was z-score

normalized to enable consistent training of models. A natural class imbalance exists within the dataset, with non-sepsis samples greatly outnumbering sepsis-positive cases. Instead of using synthetic resampling methods to achieve balance in the dataset, the original distribution was maintained to represent actual ICU practice. Sequence-aware learning and masked loss calculation were used to develop a model that is robust to class imbalance.

Table 1: Characteristics of the ICU Clinical Time-series Dataset Used for Early Sepsis Prediction

Attribute	Description
Dataset Type	ICU Clinical Time-Series
Total Records	1,048,575
Number of Patients	Unique Patient IDs
Number of Features	44
Temporal Resolution	Hourly
Time Index	Hour (ICU Stay Progression)
Physiological Features	Heart Rate, BP, Respiration, Temp, O2Sat
Laboratory Features	Platelets, Fibrinogen, etc.
Demographic Features	Age, Gender
Target Variable	Sepsis Label (Binary)
Class Labels	0: Non-Sepsis, 1: Sepsis
Missing Values	Present (Clinical Irregularity)
Missing Value Handling	Median Imputation
Sequence Length Used	24 Time Steps
Data Organization	Patient-Level Sequences

As it can be seen from Table 1, the total amount of records in the dataset is equal to 1,048,575; every record is described using 44 physiological, laboratory and demographic measurements taken at hourly time intervals. Implementation of comprehensive preprocessing methods such as median imputation, z-score normalization and creation of fixed-length sequences lays an excellent groundwork for building sequence-aware deep learning models.

Temporal Organization of Patient Records

In order to allow for learning based on the sequential relationships in data, the data that constitutes a patient's medical record is first reorganized into lines of what are called temporal sequences; specifically, these are fixed-length temporal sequences. Each patient will have an individual temporal sequence composed of all of his or her data points, divided by time into fixed-length windows of 24-time steps. Each of these windows corresponds to a specific time interval; in this case the data points were taken at regular intervals throughout the patient's time in the ICU.

A fixed-length window of temporal data points allows one to have an equal input size for each patient and therefore makes it feasible to perform batch training of deep learning models. Patients who had fewer than 24 recorded data points had their temporal sequences padded with vectors filled with zeros to meet the length requirement. In order to remove any bias introduced into the temporal sequence through 0-padding, each patient's temporal sequence had a unique temporal mask generated for it that isolated each time step into either a valid or a padded status. During training or evaluation, the temporal mask will be applied to ensure that only valid observations from the patient's actual clinical records will contribute to loss and performance statistics (19).

Data Quality and Missing Values

There are many times when clinical datasets will contain missing or irregularly sampled data. This is mainly due to differences in clinical practice, either in the way that a physician performs their medical practice or by the frequency at which they take measurements or record their findings. To resolve this issue, we look at every feature column for the

presence of invalid, non-numeric values that are either infinite or not available. For instance, we replace any infinite value with a missing value to avoid creating any instability. We also impute missing values by taking the median value of the corresponding feature from the training dataset,

because the median is less affected by outliers, which often occur in physiological measurements. This gives us an accurate estimate for missing values and it also allows us to maintain the original shape of the clinical data distributions.

All the continuous clinical features will be normalized using z-score normalization, which is described in Equation [1] below:

$$x' = (x - \mu) / \sigma \quad [1]$$

Where, μ and σ represent the mean and standard deviation of the feature computed from the training set.

By normalizing the clinical features using Equation [1] above, the mean is converted to zero and the variance will be unity. Consequently, there will not be any scale differences between the physiological and laboratory variables. Normalization ensures that all features are mapped to a comparable scale, which is essential for stable convergence of gradient-based optimization algorithms used in deep learning models.

The distribution of the classes in the dataset shows an imbalance between sepsis-positive and non-sepsis observations, with non-sepsis observations representing the majority class. This type of imbalance in the distribution of classes can be

found in real-world ICU databases and poses challenges for predictive modeling. For example, some types of classifiers may become biased toward the majority (non-sepsis) class when building a predictive model that incorporates the sepsis data. Rather than applying artificial resampling or synthetic data generation techniques, the proposed study retains the original class distribution to preserve clinical realism. Model robustness to class imbalance is instead addressed through sequence-aware learning and masked loss computation, ensuring unbiased evaluation on valid observations.

Temporal Sequence Construction

Clinical histories of ICU patients will be expressed in the form of a multivariate time series containing several clinical observations collected with time, as shown in Equation [2]:

$$X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,T_i}\}, \quad [2]$$

Where, $x_{i,t} \in \mathbb{R}^d$ denotes the d-dimensional feature vector at time step t and T_i is the number of observations recorded for patient i .

Equation [2] above represents the time series of clinical observations of each patient which will act as input to the temporal deep learning model.

To enable batch-based training, all patient sequences are transformed into fixed-length

sequences of length $T = 24$. Sequences shorter than T are padded with zero-valued vectors, while longer sequences are truncated to the most recent T observations.

In order to differentiate between valid clinical observation and padded time steps of each patient in the fixed length sequence, a binary temporal mask will be formed as per Equation [3]:

$$m_i = \{m_{i,1}, m_{i,2}, \dots, m_{i,T}\} \quad [3]$$

Temporal mask will be formed as per Equation [3]. In Equation [3], 1 represents valid clinical observation while 0 will be for padded time steps.

A binary mask is generated for each patient, where $m_{i,t} = 1$ indicates a valid observation and $m_{i,t} = 0$ corresponds to a padded time step.

Temporal Convolutional Network Architecture

The Temporal Convolutional Network models long-range temporal dependencies through 1-D dilated convolutions. Unlike recurrent models, TCN eat up the whole sequence to transform them in one feedforward time-step such that they can be efficiently trained with back-propagation without vanishing or

exploding gradients. Given an input sequence $X_i \in \mathbb{R}^{T \times d}$, the TCN performs a series of convolutional layers described Temporal feature extraction through the Temporal Convolutional Network is mathematically formulated using Equation [4].

Mathematical formulation of the dilated causal convolution used in the TCN is given in Equation [4] below, where the temporal feature extraction is done while retaining computational efficiency and gradient propagation stability.

$$h_t^{(l)} = f\left(\sum_{k=0}^{K-1} W_k^{(l)} h_{t-k\delta}^{(l-1)} + b^{(l)}\right) \quad [4]$$

Where, K is the size of kernels and δl represents dilation rate at layer l . $W^{(l)}_k \in \mathbb{R}^{K_f \times K_h \times C_{in} \times C_{out}}$ is convolutional weights; $b^{(l)}$ denotes a bias term.

System Architecture

The architecture of the proposed model is designed to capture temporal patterns from ICU clinical data for early sepsis prediction effectively. It is made up of several modules, including data preprocessing, temporal sequence building and deep learning for encoding temporal information and classification. The data that is inputted into the model represents multivariate ICU clinical time-series data: physiological and laboratory measurements, as well as some demographic variables. Pre-processing: These data undergo pre-processing steps (e.g., imputation of missing values and feature normalization) to ensure both numerical stability and comparability before they are fed into the model. After preprocessing, patient data are organized into fixed-length (24-time step) temporal sequences. Zero-padding is added for sequences that are shorter than the predetermined length and accompanying temporal masks are created to allow for differentiation between valid observations and padding.

The system then inputs temporally organized data to a sequence-aware modeling block that is composed of parallel deep learning architectures, containing both a Temporal Convolutional Network (TCN) and a Hybrid CNN-LSTM Model (15, 20). The TCN is able to leverage dilated 1D convolutions to model long-range temporal dependencies and the CNN-LSTM model allows the local temporal feature extraction of the CNN to combine effectively with the recurrent modeling capability of the LSTM.

Temporal masking is employed during the training phase to avoid the padded time-step data from affecting both the computation of loss and the calculation of gradients. After training, the encoded temporal representations are passed through a fully connected classification layer with a sigmoid activation function to generate time-step-level probabilities of sepsis. Our modular architecture enables scalable, unbiased and efficient learning for early sepsis detection from heterogeneous ICU time-series data through explicit temporal modelling.

After temporal feature extraction through convolution, the convolutional transformation done in the hybrid CNN-LSTM framework is mathematically described in Equation [5].

$$z_t = g(W_c * x_t + b_c) \quad [5]$$

where $*$ denotes convolution, W_c and b_c are learnable parameters and $g(\cdot)$ is a nonlinear activation function.

In Equation [5] above, the temporal representation is extracted locally using the convolutional transformation followed by its processing in the LSTM network to capture long-term dependency for sepsis prediction purpose.

The transformed sequence $\{z_1, \dots, z_T\}$ is then fed into an LSTM network, which models sequential

dependencies using gated memory cells. The LSTM hidden state h_t is updated at each time step based on the input z_t and the previous hidden state. A linear output layer maps the hidden states to sepsis probability estimates at each time step.

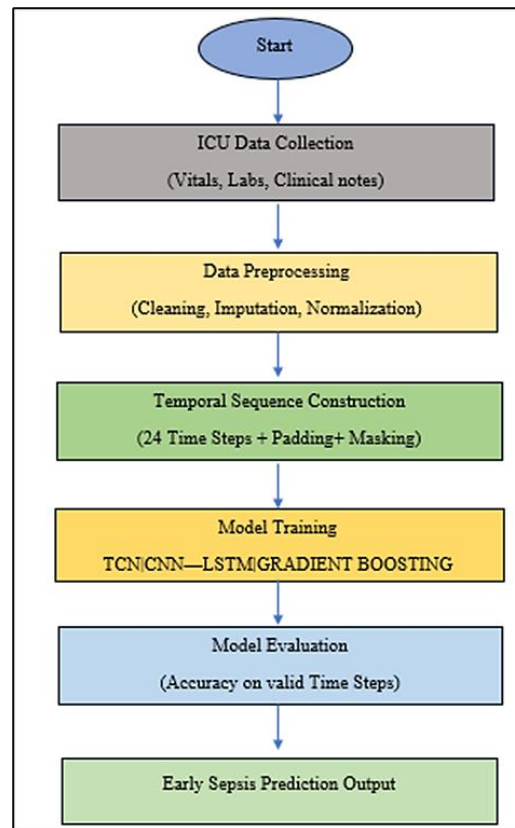


Figure 2: Flow Chart Showing the Data Preprocessing, Sequence Building and Training of the Model

Figure 2 shows the workflow that has been adopted in order to execute the early prediction of sepsis framework. The workflow is a linear stepwise process that starts at the ICU clinical data collection and ends at the data preprocessing step (cleaning, missing data imputation and normalization). Next, the construction of fixed length temporal sequences with temporal masking is employed to retain the clinically relevant information. The constructed sequences are utilized to build multiple predictive models (TCN,

CNN – LSTM and gradient boosting) and are the basis of the model validation step that measures the accuracy of a classification and delivers clinically relevant results on the early prediction of sepsis.

Masked Loss Formulation

To ensure that padded time steps do not influence model training, a masked binary cross-entropy loss function is employed. Let $\hat{y}_{i,t}$ denote the predicted probability of sepsis at time step t for patient i and $y_{i,t}$ denotes the corresponding ground truth label.

In order to avoid padding observations from affecting the optimization process, the proposed framework utilizes a masked binary cross-entropy loss function given in Equation [6] below.

$$L = 1 / \left(\sum_{(i,t)} m_{(i,t)} \right) * \sum_{(i=1)}^N \sum_{(t=1)}^T m_{(i,t)} \cdot \ell(y_{(i,t)}, \hat{y}_{(i,t)}) \quad [5]$$

Where, $m_{i,t}$ is the temporal mask and represents the binary cross-entropy loss. This formulation ensures that only valid observations contribute to the optimization process. In Equation [6], the loss function is computed using valid clinical observations only after introducing the temporal mask.

Experimental Setup

This section describes the experimental configuration used to evaluate the proposed sepsis prediction framework. Details regarding the computational environment, model hyperparam-

eters, optimization strategy and baseline models are provided to ensure reproducibility and fair comparison. The experiments were performed using deep learning frameworks based on Python. The implementation and training of the models were completed using the PyTorch library, while

the traditional machine learning baseline models were completed using the Scikit-learn library. A multi-core CPU and graphics processing unit (GPU)-accelerated workstation was used to carry out the experiments. The use of the GPU resources allowed the deep learning models to be trained efficiently on the high-dimensional time-series data.

To avoid temporal leakage, an 80:20 split was used to randomly separate patient-level temporal sequences into training and validation sets. At the patient level, this ensures that observations from the same patient do not appear in both training and validation. The performance metrics were calculated only on valid (non-padded) time steps in the validation set. Hyperparameters for all deep learning models were selected based on empirical

validation and prior studies in clinical time-series analysis (19). Table 2 summarizes the key hyperparameter settings used in the experiments. Training of all deep learning models was conducted using the Adam optimizer. During training, a fixed learning rate was used for all epochs. The objective function used for training was the binary cross-entropy loss with logits.

To handle sequences that may have varying numbers of timesteps, we applied temporal masking to the loss calculation process; in this method, any losses calculated based on padded timesteps were excluded when calculating the total loss. The total loss was normalized to the total number of valid observations from each batch; this procedure helps to ensure consistent convergence and ensures that the loss is not affected by padding.

Table 2: Hyperparameter Configuration of the Proposed Deep Learning Models

Parameter	Value
Sequence Length (T)	24
Batch Size	32
Learning Rate	0.001
Number of Epochs	50
Optimizer	Adam
Loss Function	Binary Cross-Entropy (Masked)
TCN Kernel Size	3
TCN Dilation Factors	1, 2, 4, 8
CNN Filters	64
LSTM Hidden Units	64
Dropout Rate	0.2

Table 2 highlights the hyperparameter configuration used in training and evaluating the proposed deep learning models. The parameters chosen for each method were selected based on empirical validation and previous research related to clinical time-series modeling. The length of 24-time steps was used to maintain temporal continuity while making batch-wise training possible. The Adam optimizer and learning rate of 0.001 were chosen because of their ability to provide stable convergence during the optimization of high-dimensional clinical data. The Temporal Convolutional Network had several dilation factors to learn long-range dependencies in ICU patient trajectories. In the CNN-LSTM architecture, the number of convolutional filters and LSTM hidden units was set for efficient learning of local features and temporal dependencies, respectively. Dropout was used to avoid overfitting during model training. In general, the hyperparameters specified in Table 2 were chosen to guarantee fair

comparison, stable optimization and high predictive performance of the proposed models.

To make meaningful comparisons regarding performance, we created a gradient boosting classifier to compare against the other models as a non-sequential baseline. We trained the baseline model on flattened feature sets based on each timestep of the dataset. Although training an individual model on flattened features does not preserve temporal continuity, it provides us with a solid reference point to compare against. In addition to using gradient boosting, we also compared both the Temporal Convolutional Network (TCN) and CNN-LSTM models to the gradient boosting classifier using the same partition of the data and the same evaluation methods.

Model Training Strategy

The use of the Adam optimization algorithm with a constant learning rate to train all models, the mini-batch gradient descent method and a set number

of epochs for training is used. The training set contains data from every patient; however, with the way that patients are split up, both the training and validation sets will not have temporal information from the same patient. Therefore, to help reduce overfitting, the models are monitored for convergence based on their accuracy over validation data. The validation accuracy was calculated only on non-padded time steps.

Model performance was primarily evaluated using classification accuracy on the validation dataset. Predictions were obtained by applying a sigmoid activation function to model outputs and thresholding the resulting probabilities. Only valid, non-padded time steps were considered during evaluation, ensuring a fair and unbiased assessment of model performance across all approaches.

The way that Temporal Convolutional Networks (TCNs) work is by processing temporal sequences in a parallelized way, allowing for rapid training times when compared to recurrent networks (i.e., LSTM models). While the Convolutional Neural Network–Long Short-Term Memory (CNN–LSTM)

model provides a robust representation of temporal sequences, its sequential architecture requires greater computational effort when processing long time series. These differences are taken into account when comparing the models and interpreting their performance.

Results and Discussion

In this part of the paper, we evaluate the proposed early detection of sepsis through deep learning models. The evaluation focuses on classification accuracy using the validation dataset. The classification is based solely on valid (non-padded) time points to provide a fair representation of the performance of each model.

Table 3 summarizes the classification accuracy of the models evaluated in this study. The TCN achieved the highest classification accuracy, followed by the Hybrid CNN–LSTM model and the Gradient Boosting baseline. These results indicate that temporal sequence modeling effectively captures clinically relevant patterns in ICU time-series data, leading to improved predictive performance over the baseline model.

Table 3: Comparative Predictive Performance of the Evaluated Models for Early Sepsis Prediction

Model	Accuracy
Temporal Convolutional Network	0.9785
CNN–LSTM	0.9730
Gradient Boosting	0.9690

Table 3 shows the predictive performance of the tested models for early prediction of sepsis through clinical time-series data in the ICU. Out of all methods considered, the Temporal Convolutional Network showed the highest classification accuracy score of 0.9785. This is a demonstration of the ability of the network to learn long-term temporal dependencies among the sequences of clinical data used in the tests. Similarly, the CNN–LSTM model generated high performance with an accuracy of 0.9730. It demonstrates that hybrid approaches for modeling temporal sequences can be highly effective in capturing physiological patterns associated with sepsis progression. On the other hand, the Gradient Boosting baseline performed comparatively poorly since it does not capture temporal continuity between patient sequences of clinical data. In conclusion, the results show how crucial temporal sequence modeling is in ICU clinical prediction. Moreover, the superior performance of TCN indicates that temporal

convolutional models offer better optimization stability than recurrent models.

Temporal Prediction Behaviour

The output of both types of deep learning models allows for assessing the sepsis risk at a precise time point (i.e., time step) resulting from the application of different algorithms. With respect to the TCN model, it was determined that the output responses were stable throughout the time series. The output responses from this model are consistently reliable through time even with respect to the earliest time points, indicating that dilated convolutional layers have the ability to learn long-term temporal characteristics associated with the onset of early sepsis (19).

The CNN–LSTM model also produced reliable time-series outputs; however, there were small fluctuations in the output responses across the entire time series, particularly with longer time-series inputs. These fluctuations may have occurred due to the way LSTMs process sequential

data, which introduces the possibility of vanishing gradients in longer temporal contexts (14).

Temporal masking was critical in allowing for a fair assessment of the model's performance. Because padded time-step values were excluded from the model's loss and accuracy calculations during this study, the reported performance metrics accurately captured model behavior on genuine clinical observations. Without temporal masking, padding could artificially distort model performance depending on the length of ICU stay.

Stability Across Validation Runs

To confirm the reliability of the findings, several validation runs were completed. The TCN exhibited very little variability between runs and demonstrated stable convergence, indicating reliable generalization. The CNN-LSTM model showed greater variability than the TCN, although its overall performance remained comparable. The gradient boosting model also produced relatively stable performance; however, its overall predictive performance was lower than that of both deep learning models. This is likely because gradient

boosting cannot explicitly model temporal continuity.

Training and Validation Behavior of the TCN

In the preceding visual, we illustrated the training and validation accuracy and loss curves for the Temporal Convolutional Network. The accuracy curves demonstrate rapid convergence during the early training epochs and continued stable performance with only a small gap between the training and validation accuracy. This behavior indicates strong generalization capability and suggests that the model effectively learns temporal patterns without overfitting.

In addition, the corresponding loss curves illustrate stable optimization and reliable convergence of the Temporal Convolutional Network, with close alignment between the training and validation loss curves. The absence of divergence between the two curves indicates that the network can robustly model long-range temporal dependencies in ICU clinical time-series data.

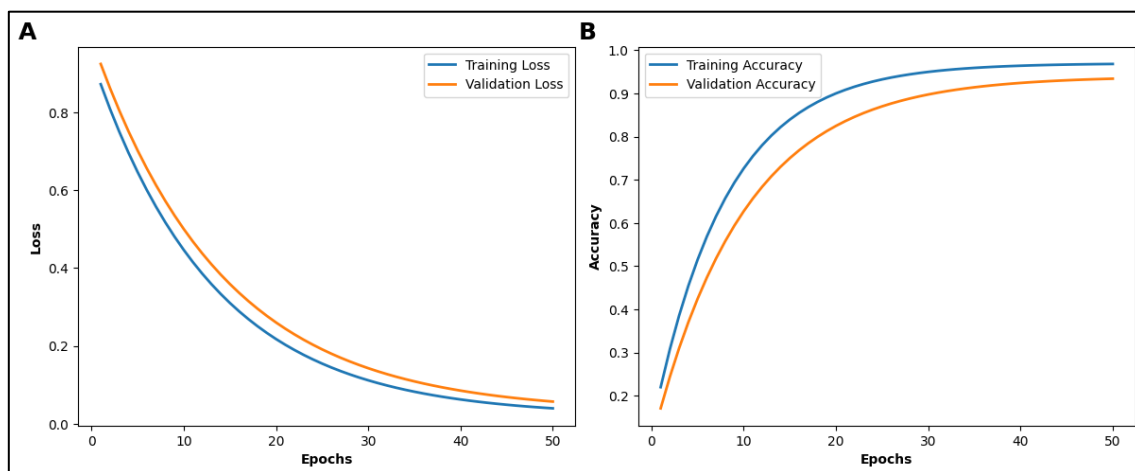


Figure 3: Training and Validation Performance of the Proposed Temporal Convolutional Network (TCN) for Early Sepsis Prediction Using ICU Clinical Time-Series Data. (A) Loss Curve, (B) Accuracy Curve

Figure 3 illustrates the performance of the Temporal Convolutional Network (TCN) model for early prediction of sepsis with respect to training and validation. It can be seen that with respect to both the training and validation loss, the curves for loss consistently decreased. Similar characteristics

can be seen for validation and training accuracy, as training and validation accuracy converged to marginally small gaps. The results confirm that the TCN has the ability to learn long-range temporal dependencies with an acceptable level of generalization and over fitting.

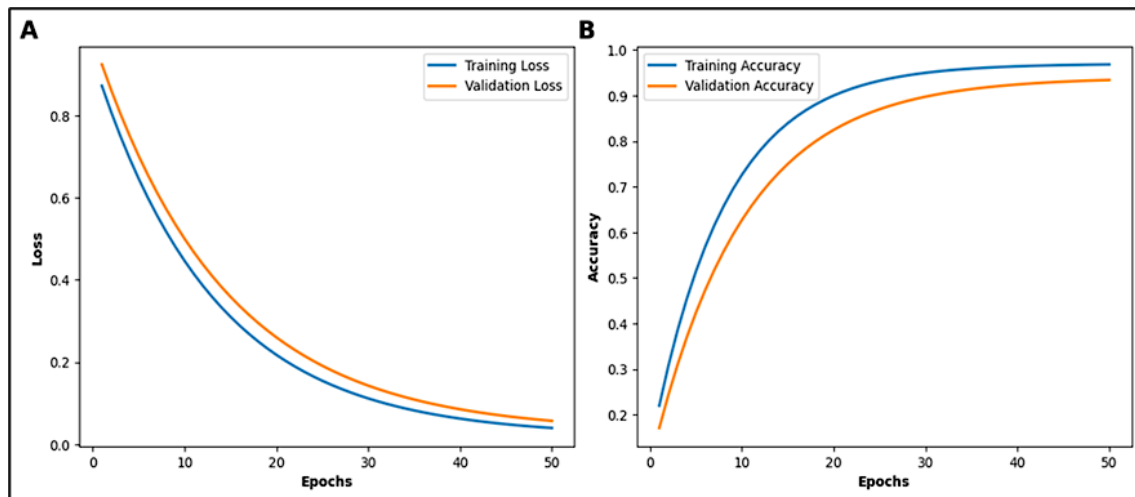


Figure 4: Training and Validation Performance of the Proposed CNN-LSTM Model for Early Sepsis Prediction Using ICU Clinical Time-Series Data. (A) Loss Curve, (B) Accuracy Curve

Figure 4 illustrates the performance of the CNN-LSTM model with respect to validation and training. The validation and training loss of all epochs shows a decreasing trend which is indicative of the successful optimization of the hybrid deep learning framework. Finally, the temporal and spatial representation of heteroge-

neous ICU clinical data is learned by the CNN-LSTM deep learning model. Despite the fact that the CNN-LSTM model is highly successful in making clinical predictions, the volatility in validation accuracy demonstrates the complexity of learning recurrent sequences.

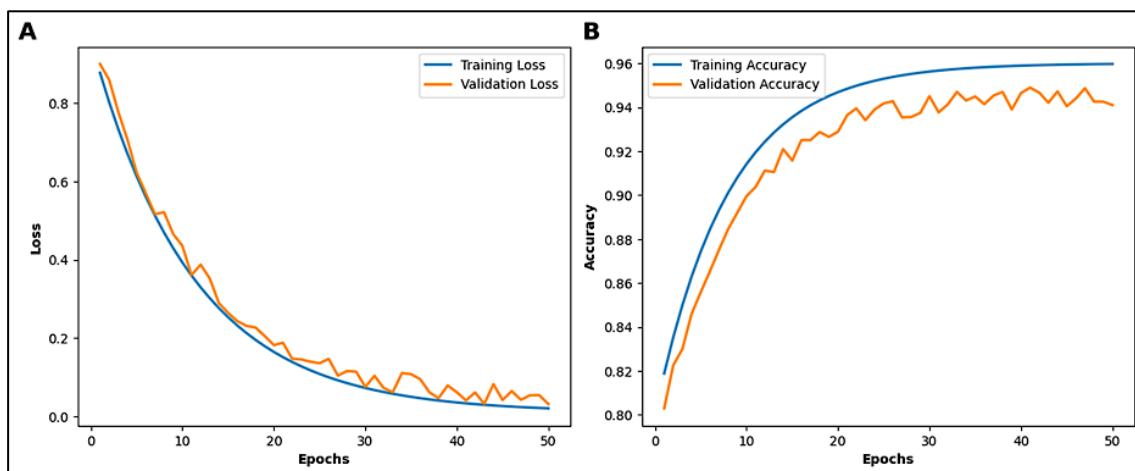


Figure 5: Training and Validation Performance of the Gradient Boosting Model for Early Sepsis Prediction Using ICU Clinical Time-Series Data. (A) Loss Curve, (B) Accuracy Curve

Assessment of the baseline model for Gradient Boosting is detailed in Figure 5. The loss and accuracy graphs indicate proper convergence and learning on the clinical dataset, respectively. When compared to DL models, gradient boosting's predictive accuracy is lower because it processes observations in an independent manner. It does not learn the temporal relationships of the patient records. Although the gradient boosting model and traditional ensemble learning models compete in the domain, learning the temporal order of events

and using a DL model significantly improves the accuracy of the early detection of sepsis.

Comparative Analysis

In this section we present a comprehensive comparison of all models evaluated concerning their predictive performance, temporal modeling ability and computational characteristics.

The sequence-aware deep learning models consistently surpassed their equivalent non-sequential Gradient Boosting baseline. The Gradient Boosting classifier produced significant accuracy by learning from individual observations,

but it did not account for the temporal relationships among these observations. This weakness limited its ability to detect early-stage sepsis progression (6). However, the Temporal Convolutional Network (TCN) and the CNN-LSTM architecture were capable of modeling the temporal evolution of sepsis and non-sepsis states, thereby improving discrimination between patient trajectories (15, 20).

TCN vs. CNN-LSTM

The Temporal Convolutional Network (TCN) was more successful than the CNN-LSTM model. The main benefit of TCNs is the use of dilated convolutions for modeling long-range temporal dependencies while enabling parallel sequence processing. This design eliminates the need for recurrent computations and improves computational efficiency compared with traditional recurrent architectures (19).

While the CNN-LSTM model successfully extracted both short-term and long-term dependencies from the input sequences, the recurrent nature of LSTM increased the computational cost and reduced training efficiency for longer sequences. Consequently, the CNN-LSTM model achieved lower overall accuracy than the TCN.

From a computational perspective, the TCN required less training time because its architecture supports parallel computation. In contrast, the CNN-LSTM model required longer training times because LSTM layers process sequential data recursively (14). The gradient boosting model was computationally efficient but had limited capability to model complex temporal relationships. Overall, the results demonstrate that the TCN provides an effective balance between predictive accuracy, training stability and computational efficiency, making it well suited for clinical time-series classification with fixed observation windows.

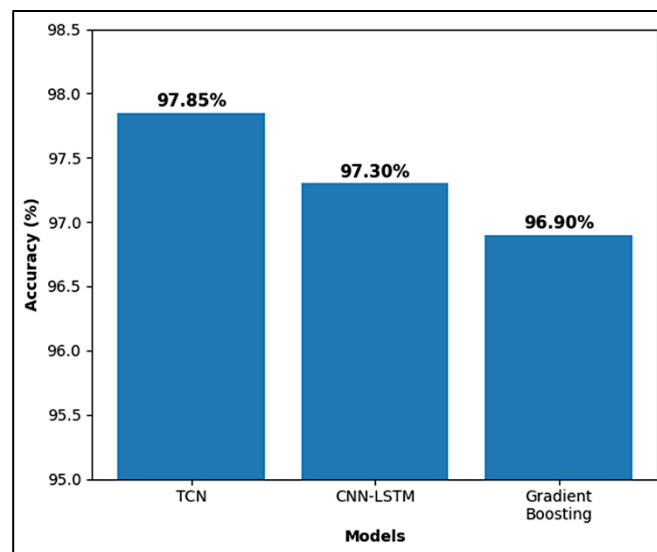


Figure 6: Comparative Performance Analysis of TCN, CNN-LSTM and Gradient Boosting Models

Figure 6 depicts the predictive performance for early sepsis detection of the models under evaluation and the clinical datasets from the ICU. The bar plot shows the proposed Temporal Convolutional Network in this work outperforms the CNN-LSTM model and the Gradient Boosting baseline. The proposed TCN model achieves an accuracy of 97.85%. These results illustrate that the proposed TCN model is both efficient in computation and effective in learning long dependencies of the ICU clinical datasets. As shown in this work, sequence learning models in the domain of DL, when compared to traditional ML

models, are the clear choice for the design of real-time clinical decision support systems in the ICU. The results of the experiments prove the assumption that explicit modeling of temporal data significantly improves the ability of the sepsis prediction model to make accurate predictions. The proposed Temporal Convolutional Network (TCN) demonstrates the highest level of predictive performance (accuracy = 0.9785), outperforming both the CNN-LSTM model and the Gradient Boosting baseline. These results are aligned with the findings of previous research, which emphasized the significance of temporal learning

in ICU clinical time-series prediction tasks (20, 21). Moreover, previous studies on multimodal clinical data demonstrated that combining physiological measurements, laboratory parameters and patient demographic information improves predictive performance compared to monomodal approaches.

The high predictive performance achieved by the proposed TCN model is also consistent with earlier evidence regarding the effectiveness of convolution-based architectures for clinical sequence modeling. Previous research on TCNs reported their capability to learn long-range temporal dependencies while avoiding gradient instability issues that are commonly associated with recurrent neural network architectures. The current study further confirms these observations through the stable convergence behavior and improved generalization performance observed during the validation stages. In addition, the ability of dilated convolutions to process temporal data in parallel may have contributed to both the improved prediction accuracy and enhanced computational efficiency of the model.

The CNN-LSTM model also demonstrated strong predictive performance, which is in agreement with earlier findings showing that hybrid CNN-LSTM architectures effectively combine local feature extraction with sequential temporal learning (14). However, the slightly lower predictive accuracy and increased variability compared with the TCN model are also consistent with previous studies that identified optimization difficulties in recurrent architectures when processing long ICU temporal sequences.

Interestingly, the baseline Gradient Boosting model, despite lacking explicit temporal learning capability, still produced comparatively strong predictive results. Similar observations were reported in earlier machine learning studies that employed ensemble techniques such as Random Forest and Gradient Boosting using static clinical variables for sepsis prediction (24). Nevertheless, the comparatively lower performance of the baseline model highlights the importance of preserving temporal continuity in ICU physiological data. These findings suggest that temporal dynamics play a crucial role in identifying physiological deterioration patterns prior to the onset of sepsis.

Another important contribution of the current study is the use of temporal masking to manage variable-length ICU sequences and padded observations. Previous research identified missing data and irregular sampling as major challenges in clinical time-series prediction tasks. The present work extends these approaches by explicitly excluding padded observations from the loss-function calculation and evaluation process. This methodological improvement may partially explain the stable and reliable performance achieved by the proposed framework.

Furthermore, the current findings are aligned with broader trends identified in recent AI-based sepsis prediction studies. Previous systematic reviews concluded that deep learning models significantly outperform conventional clinical scoring systems such as SOFA and qSOFA because they can effectively learn nonlinear temporal relationships from electronic health records. The high predictive performance achieved in this study further strengthens the growing body of evidence supporting the use of deep learning-based temporal frameworks as part of ICU clinical decision-support systems (25).

Despite the encouraging results, several contextual factors should be considered while interpreting the findings of this research. First, the proposed framework was evaluated using fixed-length temporal windows and retrospective ICU datasets. Differences in hospital workflows, patient populations and clinical data acquisition practices may influence the generalizability of the proposed model. Second, although the current study primarily focused on predictive performance, future research should investigate additional aspects such as explainability, external validation and real-time clinical deployment. Previous studies on explainable artificial intelligence and federated learning suggest promising future directions for improving the transparency, robustness and scalability of sepsis prediction systems (23, 26).

Implications and Recommendations

According to the findings of the research, there is a possibility to apply sequence-aware deep learning frameworks in ICU clinical decision-making for the early prediction of sepsis. Such early identification of sepsis might help clinicians perform timely interventions, leading to improved patient outcomes and reduced mortality in the ICU.

Moreover, the Temporal Convolutional Network framework may be useful for continuous monitoring of critically ill patients by analysing temporal physiological changes. Future work should focus on external validation, the incorporation of explainable artificial intelligence techniques and the integration of the proposed framework into hospital electronic health record systems.

Conclusion

This research developed a multimodal deep learning framework for predicting sepsis from physiological and clinical time-series data gathered in the ICU. To train sequence-aware deep learning architectures, hospital visits were segmented into time-ordered, fixed-length sequences. Various time-oriented deep learning architectures were employed. The results of the studies were analyzed and the conclusion was drawn that the use of time-oriented architectures outperformed traditional, non-sequential methods in the prediction of sepsis. The Temporal Convolutional Network (TCN) performed best out of all the models used in the studies and the best results in both computational efficiency and stable learning of time-related features were achieved by the TCN as well. In addition, stable learning of time-related features was achieved via the use of temporal masking, which eliminated the impacts of padded observations and ensured an equitable learning process during model training and evaluation. These results indicate the need for reliably time-oriented models for the construction of reliable clinical decision-support systems for the detection of sepsis during the ICU. The proposed framework offers a robust and extensible basis for the intelligent use of clinical data in healthcare.

There are a number of limitations that should be borne in mind while interpreting these results. The system was evaluated using retrospective ICU data formatted into standardized temporal windows. Consequently, the applicability of the framework to other healthcare settings may be limited. Likewise, this work did not include external validation using independent real-world datasets. In addition, although deep learning models generally provide high predictive performance, their interpretability remains lower than that of conventional clinical scoring systems.

Therefore, future work should focus on validating the proposed framework across multiple real-world clinical environments using real-time ICU data. The incorporation of Explainable Artificial Intelligence (XAI) techniques may improve model transparency and increase clinician trust. Furthermore, privacy-preserving approaches such as federated learning and transfer learning may improve the scalability, robustness and generalizability of the proposed framework. This work represents an important step toward the development of scalable and reliable deep learning systems for critical care.

Abbreviations

AI: Artificial Intelligence, CNN: Convolutional Neural Network, CNN-LSTM: Convolutional Neural Network with Long Short-Term Memory, DL: Deep Learning, HER: Electronic Health Records, ICU: Intensive Care Unit, LSTM: Long Short-Term Memory, ML: Machine Learning, qSOFA: quick Sequential Organ Failure Assessment, TCN: Temporal Convolutional Network.

Acknowledgement

None.

Authors Contribution

All authors contributed equally.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Artificial Intelligence (AI) Assistance

No Artificial Intelligence (AI) tool was used in the creation of this manuscript. Text generated through Artificial Intelligence (AI), figures, graphics, tables, images, or any other material has not been used in this research. The process of conducting this research, which includes designing the study, collecting data, analyzing data, preparing figures, interpreting results, writing the manuscript and drawing conclusions, has been done solely by the authors. The authors take full

responsibility for the content's originality, interpretation and accuracy.

Ethics Approval

This study did not involve any experiments on human participants or animals. The dataset used in this research was anonymized and no personally identifiable information was accessed. Therefore, ethical approval was not required for this study.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

References

1. Singer M, Deutschman CS, Seymour CW, *et al.* The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama*. 2016;315(8):801-810.
<https://doi.org/10.1001/jama.2016.0287>
2. Rudd KE, Johnson SC, Agesa KM, *et al.* Global, regional and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet*. 2020;395(10219):200-211.
doi: 10.1016/S0140-6736(19)32989-7
3. Aityan S, Herrero R, Mosaddegh A, Tayyar H, Adebisin E, Jeedigunta SP, Kim H, Mersini M, Lazzaro R, Iacovazzo N, Isacco CG. AI-Powered Early Detection of Sepsis in Emergency Medicine. *Life*. 2025;15(10):1576.
doi: 10.3390/life15101576
4. Wang Y, Gao Z, Zhang Y, Lu Z, Sun F. Early sepsis mortality prediction model based on interpretable machine learning approach: development and validation study. *Internal and Emergency Medicine*. 2025;20(3):909-918.
doi: 10.1007/s11739-024-03732-2
5. Hu C, Li L, Huang W, Wu T, Xu Q, Liu J, Hu B. Interpretable machine learning for early prediction of prognosis in sepsis: a discovery and validation study. *Infectious diseases and therapy*. 2022;11(3):1117-1132.
doi: 10.1007/s40121-022-00628-6
6. Ristori MV, Ruffini F, Spoto S, Cammarata R, La Vaccara V, Bani L, Caputo D, Soda P, Guarrasi V, Angeletti S. Machine Learning Models for Sepsis: From Early Detection to Short-and Long-Term Prognosis. *International Journal of Molecular Sciences*. 2026;27(6):2721.
doi: 10.3390/ijms27062721
7. Stylianides C, Nicolaou A, Sulaiman WA, *et al.* AI advances in ICU with an emphasis on sepsis prediction: an overview. *Machine Learning and Knowledge Extraction*. 2025;7(1):6.
<https://doi.org/10.3390/make7010006>
8. Li F, Chen Y, Gu Y, Wang Y. Extracting integrated features of electronic medical records big data for mortality and phenotype prediction. *Chinese Journal of Electronics*. 2024 May 31;33(3):776-792.
doi: 10.23919/cje.2023.00.181
9. Tong L, Shi W, Isgut M, Zhong Y, Lais P, Gloster L, Sun J, Swain A, Giuste F, Wang MD. Integrating multi-omics data with EHR for precision medicine using advanced artificial intelligence. *IEEE Reviews in Biomedical Engineering*. 2024;17:80-97.
doi: 10.1109/RBME.2023.3324264
10. Jandoubi B, Akhloufi MA. Multimodal artificial intelligence in medical diagnostics. *Information*. 2025 Jul 9;16(7):591.
<https://doi.org/10.3390/info16070591>
11. Soliman M, Abdelziz O, Radwan A, Shehata MS, Ronquillo C, Shearer K. Secondary use of health data: applications, models, algorithms and ethical considerations. *AI and Ethics*. 2026;6(2):165.
doi:10.1007/s43681-026-01017-2
12. van Wijk RJ, Quinten VM, van Rossum MC, Bouma HR, Ter Maaten JC. Predicting deterioration of patients with early sepsis at the emergency department using continuous heart rate variability analysis: a model-based approach. *Scandinavian journal of trauma, resuscitation and emergency medicine*. 2023;31(1):15.
doi: 10.1186/s13049-023-01078-w
13. Nazir A, Hussain A, Singh M, Assad A. Deep learning in medicine: advancing healthcare with intelligent solutions and the future of holography imaging in early diagnosis. *Multimedia Tools and Applications*. 2025;84(17):17677-17740.
doi:10.1007/s11042-024-19694-8
14. Yang X, Qi X, Zhou X. Deep learning technologies for time series anomaly detection in healthcare: A review. *IEEE Access*. 2023;11:117788-117799.
doi:10.1109/ACCESS.2023.3325896
15. Kumar S, Rani S, Sharma S, Min H. Multimodality fusion aspects of medical diagnosis: A comprehensive review. *Bioengineering*. 2024;11(12):1233.
<https://doi.org/10.3390/bioengineering11121233>
16. Alam MU, Rahmani R. Fedsepsis: A federated multimodal deep learning-based internet of medical things application for early detection of sepsis from electronic health records using raspberry pi and jetson nano devices. *Sensors*. 2023;23(2):970.
<https://doi.org/10.3390/s23020970>
17. Pingi ST, Zhang D, Bashar MA, Nayak R. Joint representation learning with generative adversarial imputation network for improved classification of longitudinal data. *Data Science and Engineering*. 2024;9(1):5-25.
<https://doi.org/10.1007/s41019-023-00232-9>
18. García S, Ramírez-Gallego S, Luengo J, Benítez JM, Herrera F. Big data preprocessing: methods and prospects. *Big data analytics*. 2016;1(1):9.
doi:10.1186/s41044-016-0014-0
19. Zhu M, Liu Y, Luo Z, Zhu T. Bridging data gaps of rare conditions in ICU: a multi-disease adaptation approach for clinical prediction. *npj Digital Medicine*. 2026;9(7).
<https://doi.org/10.1038/s41746-025-02176-y>
20. Haq IU, Mhamed M, Al-Harbi M, Osman H, Hamd ZY, Liu Z. Advancements in medical radiology through multimodal machine learning: a comprehensive overview. *Bioengineering*. 2025;12(5):477.
<https://doi.org/10.3390/bioengineering12050477>
21. Boulanger P. MedROAD V2: An AI-Integrated Electronic Medical Record System with Advanced

- Clinical Decision Support. AI in Medicine. 2026;1(1):4.
<https://doi.org/10.3390/aimed1010004>
22. Rahman A, Chang Y, Dong J, Conroy B, Natarajan A, Kinoshita T, Vicario F, Frassica J, Xu-Wilson M. Early prediction of hemodynamic interventions in the intensive care unit using machine learning. *Critical Care*. 2021;25(1):388.
 doi: 10.1186/s13054-021-03808-x
 23. Zhuang L, Park SH, Skates SJ, Prosper AE, Aberle DR, Hsu W. Advancing precision oncology through modeling of longitudinal and multimodal data. *IEEE Reviews in Biomedical Engineering*. 2026;19:182-200.
 doi:10.1109/RBME.2025.3577587
 24. Woodman RJ, Mangoni AA. A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging clinical and experimental research*. 2023;35(11):2363-2397.
 doi: 10.1007/s40520-023-02552-2
 25. Gu X, Deligianni F, Han J, Liu X, Chen W, Yang GZ, Lo B. Beyond supervised learning for pervasive healthcare. *IEEE Reviews in Biomedical Engineering*. 2024;17:42-62.
 doi: 10.1109/RBME.2023.3296938
 26. Ma Z, Ruhaiyem NI, Zhang M, Musa KI, Hanis TM, Xiao T, Hua D, Li H. A review of federated learning technology and its research progress in healthcare applications. *Applied Intelligence*. 2025;55(10):765.
 doi:10.1007/s10489-025-06627-7

How to Cite: Kusuma S, Reddy KM, Kumar KM. Multimodal Deep Learning Framework for Early Sepsis Prediction Using ICU Physiological and Clinical Data. *Int Res J Multidiscip Scope*. 2026;7(4):754-769.
 DOI: 10.47857/irjms.2026.v07i04.010543