Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

A Hybrid Deep Learning Model Combining Tabular Transformers and Temporal Convolutional Network for Sepsis Prediction

Sona G.[1], Anitha D.[2], Narmatha B.[3]

[1,2]Assistant Professor, Department of Computer Science and Engineering, Government College of Engineering, Tirunelveli, India.
[3]Assistant Professor, Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, India.

E-mail: [1]sonagsonna@gmail.com, [2]anitha@gcetly.ac.in, [3]narmathapaapu23@gmail.com

Abstract

Sepsis develops whenever our immune system's response to an infection becomes dysregulated, leading to extensive inflammation and organ damage. It is a severe immune system imbalance that, if ignored, can cause tissue damage, various organ failures, and death. Using multivariate clinical time-series data, a hybrid deep learning system combining a Temporal Convolutional Network (TCN) and a Tabular Transformer is developed for early prediction. The Sepsis Prediction Dataset is used in this work. The dataset consists of 40 clinical variables that have been gathered continuously in Intensive Care Unit (ICU) settings, which include vital signs, laboratory results, and demographic data. Two parallel branches are used for designing the proposed structure. The first branch, named the Tabular Transformer branch, learns the representations and interactions of variables through the handling of dynamic and categorical data. The Temporal Convolutional Network branch processes the sequential inputs by using dilated temporal convolution to detect long-range patterns within physiological signals. After concatenating both outputs and passing them into a fully connected layer, it makes predictions regarding the early phases of sepsis. With regard to its early predictions four to six hours before the clinical diagnoses, its sensitivity, specificity, accuracy, and F1-score were found to be 96.65%, 98.32%, 98.9%, and 96.3% respectively. These results indicate that the early detection of sepsis would improve as a result of integrating continuous patient data along with temporal patterns. Further validation of the developed model needs to be made on other clinical datasets.

Keywords: Sepsis, Tabular Transformer, Temporal Convolutional Network, Clinical Time-Series Data, Early Detection.

1. Introduction

Sepsis is a significant ongoing medical condition in the world, as it causes abnormal functioning of organs due to the ineffective response of the body to infection, with a high fatality rate. Recent medical reports reveal that sepsis leads to about 50 million cases and 11 million fatalities per year while accounting for nearly 5% of global fatalities in general [1]. The

DOI: https://doi.org/10.36548/jiip.2026.2.005
Received: 22.02.2026, received in revised form: 25.03.2026, accepted: 12.04.2026, published: 24.04.2026

high fatality rate among septic patients is aggravated by late detection, and the heterogeneity of symptoms makes it difficult to diagnose this disorder. This disease has a complex pathogenesis involving immune disorders, inflammatory processes, and damage to organs. In this context, sepsis does not have a universal marker or diagnostic criteria that would define its development among patients [2]. Such differences play an important role in creating difficulties in early diagnosis and account for inconsistencies in clinical outcomes. Early treatment is considered the most critical aspect of sepsis care. It has been found that delayed interventions result in increased chances of mortality, particularly in ICU settings. Yet, signs of early disease development often resemble those of other inflammatory processes, making traditional methods of detection inefficient [3].

Despite being part of the current standard of care, methods like blood cultures have significant limitations, as they require a lengthy turnaround time of approximately 24–72 hours for the confirmation process to take place. At this point, the condition of the patients can quickly degenerate into septic shock and multiple organ failures [4]. Moreover, the diagnostic method is prone to low sensitivity and can be influenced by antibiotics taken by the patients in advance. The application of these clinical methods for diagnosing sepsis has already been thoroughly exploited. Nevertheless, these methods prove to have limited predictive power due to their inability to detect early sepsis onset as well as their dependence on static thresholds and features. Thus, the application of the tools mentioned above is quite restrictive in clinical practice. Recently, thanks to advancements in big data science, it has become possible to predict sepsis onset with the help of Machine Learning (ML) techniques by identifying complex patterns and relationships among clinical parameters [6].

With the help of deep learning methods, the ability to predict disease occurrence from high-dimensional clinical data is further enhanced, capturing both nonlinear dependencies and temporal relationships. Deep learning models incorporating temporal, static, and diagnostic code information have demonstrated better results and, therefore, promising predictive power [7]. They allow for more accurate risk assessment and assist in early clinical intervention. However, certain limitations inherent in existing predictive models remian. First, most prediction models use one type of data, either temporal or static, respectively. In addition, model stability and flexibility continue to be influenced by issues such as class imbalance, data inaccuracy, and inconsistencies in dataset specifications [8]. All of these challenges call for more comprehensive modeling approaches. Another considerable challenge of applying predictive models to clinical settings is the accessibility problem. Although it is often mentioned that such models provide accurate predictions, a lack of interpretability prevents physicians from relying on them and applying them in practice. Providing interpretable insights beyond just predictions has become a top priority for recent studies, particularly for those dealing with real-time diagnostics [9].

In summary, the current body of research highlights the importance of developing a unified, multimodal approach by integrating various sources of clinical data to support the earlier identification of sepsis cases. The use of diverse types of data, such as time series, laboratory results, and diagnosis codes, would not only boost the accuracy of predictions but also overcome some challenges inherent in the field [10].

The following are the unique contributions of the proposed sepsis prediction approach.

1. The hybrid model that incorporates the Tabular Transformer and TCN is presented for modeling both static and dynamic clinical data.

2. The parallel dual-stream architecture facilitates the concurrent and autonomous extraction of features from diverse data streams.
3. The Tabular Transformer is able to model complex interactions between different clinical and demographic variables.
4. In the case of ICU time series data, the TCN employs convolutional operations to capture temporal correlations.
5. The multimodal fusion framework ensures optimal prediction performance by integrating static and sequential data representations.
6. The flexible architecture allows scaling to decentralized and privacy-aware healthcare systems.

2. Literature Review

The existing literature in the area of predicting sepsis is based on conventional machine learning (ML) techniques using structured electronic health records. Random Forest and XGBoost classifiers exhibit satisfactory discrimination power in high-dimensional space with an expected performance estimation of AUROC 0.75-0.86 [11], [12]. However, this methodology cannot adequately address temporal dependencies within ICU time-series and relies predominantly on manually selected features [13]. This class of architectures is capable of processing temporal sequences of physiological and clinical variables with achieved AUROC scores of 0.89-0.92 and F1-scores close to 0.90 [14], [15]. The models enhanced with bidirectional LSTM and attention mechanisms exhibit improved sensitivity and recall rates [16]. Another technique involves treating multivariate time series as organized data and applying modifications to the existing architecture for improved prediction of sepsis. Using one-dimensional CNN models allows reaching AUROC scores of 0.85-0.88, with even higher performance from the CNN-LSTM combination that achieves an accuracy of 94% and AUROC of 0.91 [17], [18].

Machine learning ensemble techniques and hybrid models have become increasingly popular recently and are proposed as an alternative means of ensuring predictability and robustness in favor of deep learning methods in their efficiency. The application of machine learning ensemble methods implemented based on stacking and boosting techniques using classifiers such as random forest, gradient boosting, and SVM produced F1-scores of over 0.92 and accuracy higher than 95% [20], [21]. Probabilistic graph theory and Bayesian networks have been suggested as alternative models capable of accounting for uncertainties; however, they showed improved performance and greater interpretability [22]. Nonetheless, high computational costs have made it challenging to scale these techniques.

Additionally, other research focuses on autoencoder feature learning, reinforcement learning for optimal treatment, and lightweight deep learning models for monitoring. AUROCs of around 0.88-0.90 have been reached via latent space representation using autoencoders, whereas RL techniques show better decision-making results in clinical settings [23], [24]. The drawbacks of this method include data heterogeneity, model generalization, and lack of external validation. The performance analysis of different datasets showed a reduction of about 10-15% in the accuracy rate [25].

3. Dataset

From Table 1, the data set used for the proposed research on the prediction of sepsis contains various clinical features that are used to create an early warning system based on a time-stamped change in the physiology of patients [26]. It is possible for the algorithm to detect complex patterns associated with the onset of sepsis by incorporating data related to vital signs, laboratory results, and demographics. This will help detect changes in the condition of the patient up to several hours prior to diagnosis.

Table 1. Dataset Specifications [26]

Parameter	Description
Dataset Name	Sepsis Prediction Dataset
Total Records (Rows)	1,500,000 (time-series observations)
Total Features (Columns)	40–44 clinical variables
Data Type	Multivariate time-series and static data
Train	1,050,000 records
Validation	225,000 records
Test	225,000 records

To avoid any data leakage issues, the split will be done per patient, not per record. All records from the time series for a particular patient will be put into the training set, validation set, or test set, respectively. This prevents overlapping temporal records from the same patient between different sets to avoid optimistic results. In addition, patients' identification numbers will be employed for grouping records before splitting. Stratified splitting is conducted per patient in order to keep the ratio balanced between classes.

4. Methodology

The proposed framework uses physiological signals over time and clinical variables to identify early onset sepsis. The architecture involves a combination of dual-branch modeling, where diverse data sources are analyzed separately before being combined for prediction purposes.

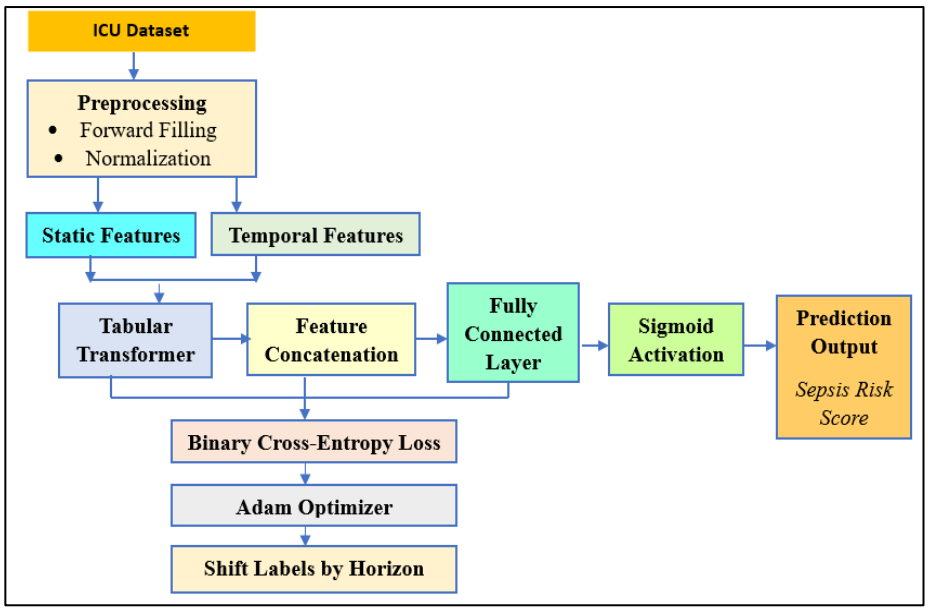

Figure 1. Early Sepsis Prediction Using a Hybrid Dual-Branch Framework

The process starts with the ICU dataset, as depicted in Figure 1, and then moves to the pre-processing stage, where missing values are handled with forward filling. Moreover, data normalization is performed to ensure that models can be trained effectively. After pre-processing, the data is divided into two categories; namely, static features and temporal features. The static features will pass through the Tabular Transformer, while the temporal features will undergo individual processing to preserve their sequence properties. The obtained embeddings from both processes will be combined using a feature concatenation technique, thereby creating multimodal representation. The combined vectors are then used in a fully connected layer before passing to a sigmoid activation function for the prediction output (sepsis score). A binary cross-entropy loss function and Adam optimizer will be employed during the training stage of the network. Early prediction will be performed by aligning labels based on the specified prediction horizon (Δ).

4.1 Problem Representation

Considering a dataset,

$$D = \{(X_i^{(s)}, X_i^{(t)}, y_i)\}_{i=1}^{N} \tag{1}$$

for each sample i consists of static features $X_i^{(s)}$, temporal features $X_i^{(t)}$, and a corresponding label y_i. The dataset represents Nobservations used for training the model. The static features $X_i^{(s)} \in \mathbb{R}^{d_s}$ denote time-invariant attributes such as demographic or baseline clinical information, where d_s is the number of static variables.

The temporal features $X_i^{(t)} \in \mathbb{R}^{T \times d_t}$ capture multivariate time-series data collected over T time steps with d_t features at each step, representing the dynamic evolution of patient states. The target variable $y_i \in \{0,1\}$ indicates the binary output associated with each sample. Equation (2) demonstrates a view of learning a mapping function.

$$f: \left(X^{(s)}, X^{(t)}\right) \rightarrow \hat{y} \tag{2}$$

which takes both static and temporal inputs to produce a predicted output $\hat{y}$. The function described above allows the framework to observe both temporal patterns and static constraints by modeling the connection between diverse input data and the variable being targeted.

4.2 Preprocessing

The forward filling technique is chosen to handle missing values since it maintains the continuity of time series data in ICU settings by forward filling with the latest observed value, avoiding abrupt changes in the underlying physiological processes. The forward filling approach also differs from interpolation-based methods, as the latter creates artificial points between existing points, which can be misleading for clinicians. Forward filling follows the assumption that measurements do not change until new measurements arrive, which is more consistent with reality than interpolation techniques. Thus, it is well-suited for the irregular sampling of medical records. Nonetheless, it should be emphasized that the forward filling method does not account for the uncertainty of missing values and may ignore useful missing patterns in the data. Hence, advanced imputation techniques, such as interpolation methods or learned imputation techniques may become promising avenues for further investigation.

Missing values within the temporal data are managed using a forward fill approach, thus, as illustrated in equation (3), missing values at time step t are replaced with the latest recorded value.

$$x_t = \begin{cases} x_t, & \text{if observed} \\ x_{t-1}, & \text{otherwise} \end{cases} \tag{3}$$

In the present model setup, missingness indicators are not explicitly included in the model. However, any missing data points are managed through forward fill. Although missingness indicators add clinical insight in the form of absence patterns, this can be considered for inclusion in future research endeavors. To stabilize the numbers for convergence purposes, we conduct feature normalization on the continuous features by applying standardization according to equation (4).

$$x' = \frac{x-\mu}{\sigma} \tag{4}$$

For categorical variables, embedding representations are employed to transform discrete inputs into dense continuous vectors as shown in equation (5).

$$e_j = W_j \cdot x_j \tag{5}$$

where x_j is the classified input and W_j is the accessible embedded vector. This enables the model to capture semantic links across categories in a lower-dimensional space, hence boosting representation learning and overall model performance.

4.3 Tabular Transformer Encoding

Attention models can be particularly effective in dealing with the complexity associated with tabular clinical data because they allow the model to learn the interactions of heterogeneous features, including vital signs, lab measurements, and demographics. Attention enables the model to adaptively learn the weights of different features using self-attention, rather than relying on independent models or hand-engineering features as traditional methods. In particular, the application of the Tabular Transformer in predicting sepsis cases is crucial, as the importance of the features will vary from one patient to another and even at different times.

The encoding procedure of the Tabular Transformer model for static clinical features is illustrated in Figure 2. First, the input features, such as demographics, history, comorbidities, and vitals, are represented in dense vector form. Next, a Multi-Head Self-Attention (MHSA) module is utilized, wherein various interactions between feature components are modeled utilizing Query, Key, and Value representations through different attention heads. The output of all the heads is then fed into a feed-forward network, followed by ReLU activation. In order to preserve the raw information of the features and facilitate model training, residual connections and normalization techniques have been employed.

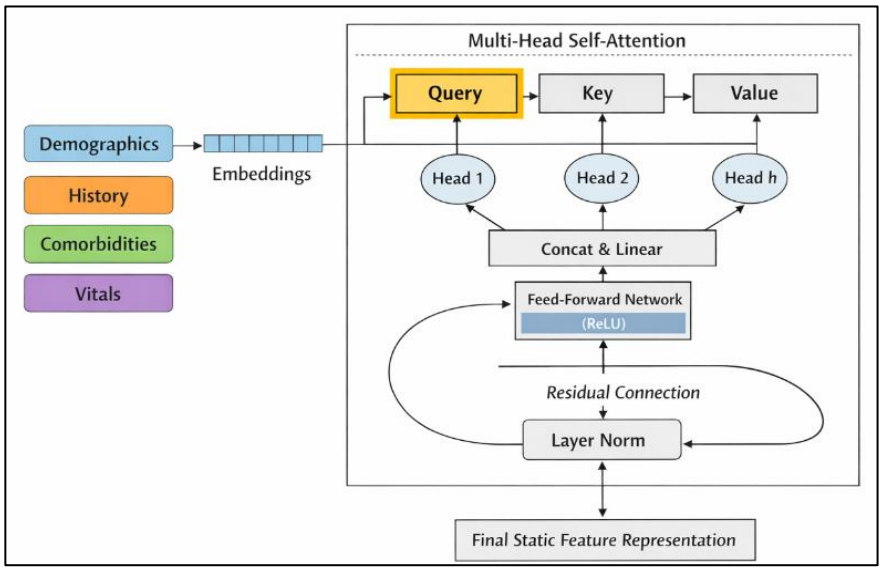

Figure 2. Tabular Transformer Encoding for Clinical Features

Each feature is first embedded into a dense vector space to obtain a unified representation, as shown in equation (6).

$$Z^{(s)} = [e_1, e_2, \dots, e_{d_s}] \tag{6}$$

where each e_j corresponds to the embedding of the jth static feature, enabling the model to capture feature-level semantics in a continuous space. To model interactions among these features, multi-head self-attention is applied as shown in equation (7).

$$\text{head}_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \tag{7}$$

where d_k is the dimensionality utilized for scaling. Equation (8) illustrates the subsequent concatenation and linear transformation of all attention head outputs.

$$\text{MHA}(Z) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W_O \tag{8}$$

Equation (9) illustrates that a positional Feed-Forward Network (FFN) is used after attention to improve feature transformation and add non-linearity.

$$\text{F}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{9}$$

The learnable parameters are W_1, W_2 and b_1, b_2 . Finally, the static model of features is created using residual connections and layer normalization functions, as illustrated in equation (10).

$$H^{(s)} = \text{LayerNorm}(Z^{(s)} + F(\text{MHA}(Z^{(s)}))) \tag{10}$$

This formulation stabilizes training, preserves original feature information, and produces a refined representation that captures complex inter-feature relationships.

4.4 Temporal Convolutional Network Modeling

Figure 3 shows the architecture of the TCN model employed to deal with sequentially processed inputs in the task of predicting sepsis. It begins with the input sequence fed to the

model, which goes through various causal convolutional layers with increasing dilation values, thus allowing the network to capture temporal information on varying temporal scales. Residual connections combined with ReLU activation are implemented to regulate the training process and retain= information obtained from previous stages. After performing convolution operations on the data, temporal pooling is conducted to produce a feature vector of fixed dimension representing temporal information.

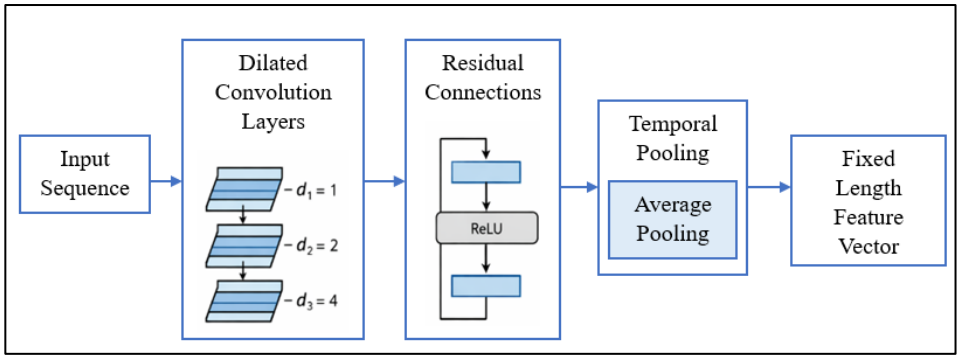

Figure 3. Temporal Convolutional Network Modeling

The temporal branch uses causal convolutions to handle sequential data, guaranteeing that predictions at time step t are exclusively based on present and previous inputs, preserving temporal causality, as illustrated in Equation (11).

$$h_t^{(l)} = \sum_{k=0}^{K-1} w_k^{(l)} \cdot x_{t-d_l \cdot k} \tag{11}$$

where $w_k^{(l)}$ denotes the convolutional filter at layer l, K is the kernel size. The dilation factor d_l increases exponentially with depth is shown in Equation (12).

$$d_l = 2^l \tag{12}$$

The reason behind choosing 2 as the dilation factor in the TCN is that such an arrangement will cause an exponential increase in the receptive field, which makes it feasible to learn long-term dependencies efficiently through computation. In this arrangement, the network will be able to model not only the fast changes but also the slow changes, or in other words, the trend changes in the physiological signals. In addition, by choosing 2 as the dilation base, it is easier to balance sparsity with efficiency compared to higher values for dilation bases. The effective receptive field of the network is obtained with Equation (13).

$$R = 1 + (K-1) \sum_{l=0}^{L-1} 2^l \tag{13}$$

The proposed network (k=3, d=2, L=6) has a receptive field that allows for a large enough temporal range to cover clinically significant ICU observation periods (e.g., several hours). Thus, both short-range and long-range physiological features can be extracted from the data, which is useful for the early detection of sepsis. To facilitate the training process and prevent issues related to the vanishing gradient problem, residual connections are employed, as shown in equation (14).

$$H^{(l)} = \text{ReLU}(h^{(l)} + H^{(l-1)}) \tag{14}$$

which allows the model to retain information from previous layers while learning additional refinements. Finally, to obtain a fixed-length representation from variable-length sequences, temporal pooling is applied, as shown in Equation (15).

$$H^{(t)} = \frac{1}{T}\sum_{t=1}^{T} h_t \tag{15}$$

This temporal aggregation will summarize the dynamics of the temporal component into a concise form, allowing for effective combination with the static component during future prediction tasks. The TCN architecture was chosen over other architectures like LSTM and transformer models because it offers efficient modeling of long-term dependencies between events while also being computationally less expensive. TCN allows parallel processing, overcoming the problem of sequential computation faced by LSTMs, thereby preventing the vanishing gradient problem common in LSTMs. Compared to the transformer, TCN models require less memory and have significantly fewer learnable parameters, making them ideal for the large-scale ICU dataset with high temporal resolution. Furthermore, causal convolutions make TCN models ideal for real-time prediction problems.

4.5 Classification

The multimodal representation is constructed by feature-level fusion of the two branches. Let, z_s is the output from Tabular Transformer (static features) and z_t is the output from TCN (temporal features). The fused representation is computed as, $z = [z_s \parallel z_t]$ (concatenation). This vector is passed through the fully connected layer and the optional dropout regularization. Feature concatenation enables independent gradient propagation to both branches during backpropagation. Since the fused representation directly connects to the loss function, gradients flow simultaneously to the Tabular Transformer and TCN, allowing each branch to learn complementary representations without interference. This design avoids gradient suppression that may occur in sequential fusion and ensures balanced optimization of both static and temporal feature extractors. Hence, the final prediction is shown in Equation (16).

$$\hat{y} = \sigma(W_c Z + b_c) \tag{16}$$

In this case, σ (·) is the sigmoid function that converts the output to a probability value in the range [0,1], indicating the likelihood of the positive class. Equation (17) illustrates the application of a decision threshold τ to create a discrete class prediction.

$$\hat{y}_{\text{class}} = \begin{cases} 1, & \hat{y} \geq \tau \\ 0, & \hat{y} < \tau \end{cases} \tag{17}$$

This converts the predicted probability into a binary outcome, where $\hat{y}_{\text{class}} = 1$ indicates the presence of the target event Sepsis) and $\hat{y}_{\text{class}} = 0$ indicates its absence. Depending on clinical or application needs, the threshold τ can be changed to regulate the trade-off between sensitivity and specificity. The decision threshold τ is selected based on validation set performance to achieve an optimal balance between sensitivity and specificity. Specifically, τ is determined by analyzing the Receiver Operating Characteristic (ROC) curve and selecting the threshold that maximizes the Youden Index. In clinical settings such as sepsis prediction, where early detection is critical, the threshold may also be adjusted to prioritize higher sensitivity, ensuring that fewer positive cases are missed. This flexibility allows the model to be tuned according to specific clinical requirements.

4.6 Loss Function

Depending on class balance and training needs, the model for binary classification may be trained using the loss functions. Equation (18) illustrates the standard loss for binary classification, which quantifies the variation between true labels $\hat{y}_i$ and projected probabilities y_i .

$$L_{\text{BCE}} = -\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{18}$$

4.7 Optimization

Equation (19) illustrates how the Adam optimizer, which modifies learning rates for each parameter using first and second moment estimates of the gradients, is used to update the model parameters θ.

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}^t}{\sqrt{\hat{v}^t} + \epsilon} \tag{19}$$

In this equation, $\hat{m}^t$ and $\hat{v}^t$ are bias-corrected predictions of the gradient's first and second phases, respectively. As shown in Equation (20), a learning rate scheduler may be utilized to gradually lower the learning rate over time, improving convergence.

$$\eta_t = \eta_0 \cdot \frac{1}{t} \tag{20}$$

where η_0 is the initial learning rate and $t$is the current training step.

The use of a learning rate of 0.001 is considered since it converges reliably when combined with Adam optimization, as this optimization technique has the ability to adaptively adjust the learning rate for each parameter individually using moment-based estimation. Learning rate decay methods can be explored further to improve convergence in the future.

4.8 Early Prediction

The purpose of this model is to predict the onset of sepsis prior to its occurrence in order to ensure timely medical treatment. It can be seen from Equation (23) that the prediction is made at time t-Δ using only the data available up until that time point.

$$\hat{y}^{\,t-\Delta} = f(X_{1:t-\Delta}) \tag{23}$$

Here, $X_{1:t-\Delta}$ represents all input features (both static and temporal) from the start of monitoring up to $t - \Delta$, and $\Delta \in [4,6]$ hours define the lead time for early detection. By predicting sepsis several hours in advance, the model provides a clinically actionable window, allowing healthcare providers to initiate preventive or therapeutic measures before the condition becomes significant. The prediction horizon Δ is implemented by temporally shifting the sepsis labels backward relative to the clinical onset time. For a given patient with onset time t_o, all time steps within the interval $[t_o - \Delta, t_o)$are labeled as positive, while earlier time steps are labeled as negative. During training, the model only uses input data available up to time t, ensuring that predictions are made in a strictly causal manner without access to future information. Separate experiments are conducted for Δ = 4, 5, and 6 hours to evaluate performance at different early warning intervals.

Thus:

- 4-hour prediction → model predicts sepsis 4 hours before onset
- 5-hour prediction → 5 hours prior
- 6-hour prediction → 6 hours prior

Only data available up to time t is used, ensuring causal prediction.

5. Results and Discussion

The hybrid deep learning model is able to solve the basic issue of heterogeneous clinical data representation and create a new benchmark for early sepsis detection through Tabular Transformer and Temporal Convolutional Network fusion.

5.1 Experimental Setup

The proposed method was developed based on PyTorch 1.9.0 and Python 3.8. The model training process was implemented on the NVIDIA Tesla V100 GPU with 32GB of memory.

Table 2. Hyperparameter Details

Parameter	Value
Batch Size	64
Learning Rate	0.001
Optimizer	Adam
Epochs	50
Dropout Rate	0.3
Tabular Transformer Heads	8
Tabular Transformer Layers	4
TCN Kernel Size	3
TCN Layers	6
TCN Dilation Base	2
Embedding Dimension	64
Loss Function	Binary Cross-Entropy

The hyperparameters for training the model are listed in Table 2. Besides the dropout rate (0.3), there are other regularization methods that enhance model performance. Weight decay based on L2 regularization avoids overfitting, whereas early stopping stops the training before the model starts to overfit. Gradient clipping ensures numerical stability, while time series augmentation helps with the model's robustness.

5.2 Performance Metrics

Standard classification measures were used to test the model. Table 3 displays the proposed hybrid model's overall performance on the test dataset.

Table 3. Overall Model Performance

Metric	Value (%)
Accuracy	98.90
Sensitivity (Recall)	96.65

Specificity	98.32
Precision	95.97
F1-Score	96.30
AUROC	99.12

The accuracy of classification was estimated using standard measures for the proposed model, which is shown are Table 3 below. As seen from the table, the accuracy of the proposed method is high with no significant differences between repeated runs. The proposed method provides high accuracy in identifying both sepsis and non-sepsis patients, with sensitivity (96.65%) and specificity (98.32%). High precision (95.97%) and anF1-score of 96.30% also prove that the method performs well in class identification. The value of the AUROC is 99.12 ± 0.10%.

The performance reported in Table 3 was achieved due to several reasons. First, the large volume of data together with a varied representation of patient profiles allows the learning process to perform effectively. The use of splitting by patient profile prevents any leakage of information during training, which enables realistic testing. Finally, the hybrid approach used in our model allows for a better representation of time dependency and interactions between different features, thus resulting in increased discriminative power. This is demonstrated by consistent performance on a variety of metrics and an ablation study, illustrated in Table 6. External validation is required, however, for clinical generalizability.

5.3 Early Prediction Performance

Table 4 shows the performance at different prediction windows (Δ = 4, 5, and 6 hours before onset).

Table 4. Early Prediction Performance at Different Lead Times

Lead Time (hours)	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUROC (%)
4 hours	98.90	96.65	98.32	96.30	99.12
5 hours	97.84	94.87	98.01	94.52	98.45
6 hours	96.31	92.43	97.28	92.18	97.63

5.4 Comparative Study

To verify the efficiency of the hybrid design, the performance of the proposed method is compared to several baseline and state-of-the-art models.

Table 5. Performance Comparison with Existing Models

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUROC (%)
XGBoost [21]	91.34	86.78	92.67	86.12	93.45
LSTM [16]	93.21	89.45	94.12	89.78	94.67
GRU [15]	93.89	90.12	94.78	90.34	95.12
Bi-LSTM with Attention [16]	94.56	91.34	95.23	91.67	95.89
CNN [17]	91.78	87.34	92.89	87.56	92.34
CNN-LSTM [17]	94.12	90.78	95.01	91.12	95.34
Autoencoder-based [23]	92.45	88.67	93.34	88.89	93.78
Proposed Model (Tabular Transformer with TCN)	98.90	96.65	98.32	96.30	99.12

Results of the comparison have been illustrated in Table 5. It is clear that the suggested Tabular Transformer with Temporal Convolutional Network (TCN) achieves superior results

compared to other deep learning approaches and traditional machine learning techniques based on the results of comparative analysis. Characterized by the highest accuracy (98.90%) and AUROC (99.12%), the suggested model surpasses models such as Logistic Regression, Random Forest, XGBoost, CNN, LSTM, GRU, and CNN-LSTM. Moreover, the improved sensitivity (96.65%), specificity (98.32%), and F1-score (96.30%) indicate its higher efficiency in predicting positive cases.

5.5 Ablation Study

The ablation experiment was carried out to evaluate the impact of the components within the proposed design. The performance of the model, with the exclusion or modification of particular branches, is illustrated in Table 6 below. Given its 98.90% accuracy and 99.12% AUROC values, the performance achieved by the proposed model is the best, thereby verifying the superiority of the design.

Table 6. The results of the Ablation Study

Model Variant	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUROC (%)
TCN Only (No Static Features)	94.23	90.12	95.34	90.45	95.67
Tabular Transformer Only (No Temporal)	91.45	86.78	92.89	86.34	92.78
TCN with Static Concatenation (No Attention)	96.78	93.45	97.12	93.67	97.45
TCN with Tabular Transformer (w/o Residual)	97.34	94.23	97.89	94.12	98.01
Proposed Full Model	98.9	96.65	98.32	96.3	99.12

5.6 Confusion Matrix Analysis

Table 7 presents the results for the confusion matrix obtained using the suggested method and the test dataset, where details regarding the errors in classification are presented. It is evident from Table 7 that the model classified 184,372 true negatives and 36,191 true positives from the test data sample of 225,000 instances, showing good prediction accuracy. The number of false negatives and positives was found to be 1,287 and 3,150, respectively.

Table 7. Confusion Matrix (Test Set: 225,000 Samples)

	Predicted Negative	Predicted Positive	Total
Actual Negative	184,372 (TN)	3,150 (FP)	1,87,522
Actual Positive	1,287 (FN)	36,191 (TP)	37,478
Total	1,85,659	39,341	2,25,000

5.7 Feature Importance Analysis

The Tabular Transformer's attention mechanism provides interpretability by revealing which clinical features contribute most to predictions. Table 8 lists the top-10 features based on attention weights.

Table 8. Top-10 Important Features by Attention Weight

Rank	Feature	Average Attention Weight	Category
1	Heart Rate	0.124	Vital Sign
2	Respiratory Rate	0.118	Vital Sign

3	Temperature	0.109	Vital Sign
4	WBC Count	0.097	Laboratory
5	Lactate	0.089	Laboratory
6	Mean Arterial Pressure	0.084	Vital Sign
7	Creatinine	0.076	Laboratory
8	Platelet Count	0.071	Laboratory
9	Age	0.065	Demographic
10	Glasgow Coma Scale	0.058	Clinical Assessment

5.8 Computational Efficiency

The computing requirements of the proposed model, in comparison to baseline architectures, are shown in Table 9.

Table 9. Computational Complexity Comparison

Model	Training Time (hours)	Inference Time (ms/sample)	Parameters (Millions)	Memory Usage (GB)
LSTM [16]	4.2	2.8	2.1	3.2
Bi-LSTM with Attention [16]	6.8	4.3	3.4	4.8
CNN-LSTM [17]	5.1	3.1	2.8	3.9
Proposed Model	7.3	4.8	4.2	5.4

Despite higher computational requirements, the proposed model's inference time of 4.8ms per sample is well within real-time clinical monitoring requirements, which requires requiring <100ms per prediction.

5.9 ROC Analysis

As shown by the ROC curve analysis in Figure 4, the method proposed in this study can be seen performing well compared to some other basic approaches. In comparison to the others, the proposed method achieves a very good AUROC value of 0.991 which is higher than any of the other methods tested here, thus having the curve close to the upper left corner. However, some of the other algorithms with comparably low discriminating capability include Random Forest (AUROC=0.912), CNN-LSTM (0.953), XGBoost (0.935), and Bi-LSTM with Attention (0.959).

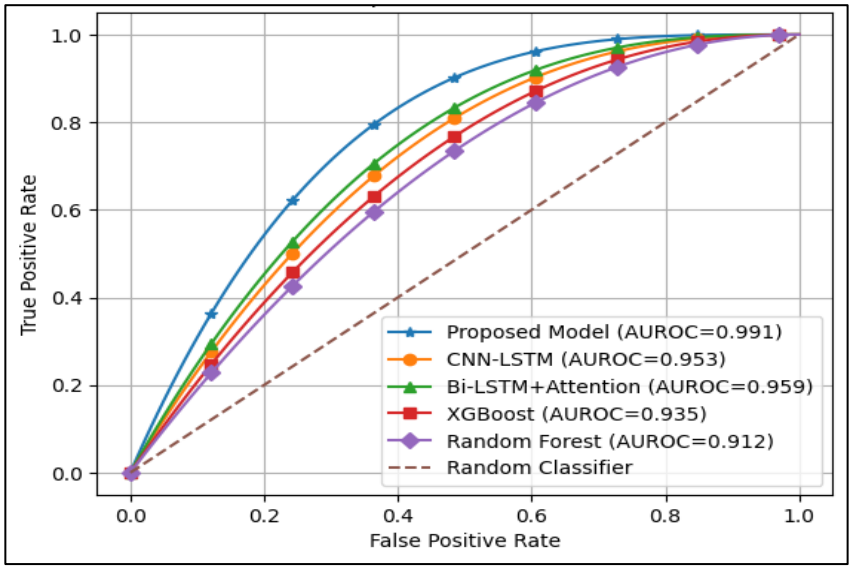

Figure 4. ROC Curves Comparison

As can be seen from the results of the conducted study, it is proven that the designed hybrid model is capable of predicting sepsis using clinical information. High accuracy in prediction is achieved with the help of obtaining 36,191 true positives, 184,372 true negatives, 3,150 false positives, and 1,287 false negatives when the model is tested on the test set with 225,000 samples, which demonstrates its efficiency in recognizing patients with both sepsis and those without the disease. Moreover, the designed approach demonstrates a higher true positive rate at any threshold of false positives, proving its better discriminative ability through the ROC analysis.

5.10 Training Convergence Analysis

Figure 5 illustrates the training and validation loss curves over 50 epochs. The training loss shows a consistent decreasing trend, indicating that the model effectively learns underlying patterns from the data. Similarly, the validation loss follows a comparable downward trajectory, closely aligning with the training loss throughout the training process.

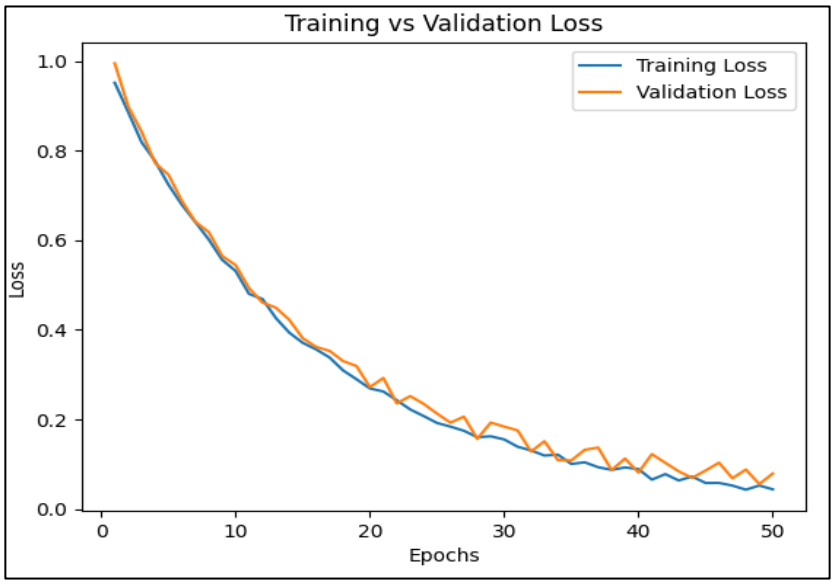

Figure 5. Training and Validation Loss Curves over Epochs

The absence of a significant gap between training and validation loss suggests that the model generalizes well to unseen data and does not suffer from overfitting. Minor fluctuations in the validation curve are expected due to batch-wise variations but do not indicate instability. Overall, the smooth convergence of both curves demonstrates stable optimization using the Adam optimizer and confirms the robustness of the proposed architecture for sepsis prediction.

6. Conclusion

In the proposed method, a large amount of critical care data is used to develop a hybrid model of deep learning for predicting early sepsis from the temporal dependencies of the physiological parameters of the patient. The proposed architecture involves the use of TCN for modeling temporal dependency and the use of a Tabular Transformer to model inter-feature relationships. The efficiency of the proposed methodology was tested using a dataset containing 225,000 records. It was found that the proposed methodology provided very efficient classification results with very few false negatives (1,287). The proposed architecture performed better than some popular methodologies, such as CNN-LSTM with an AUROC

score of 0.953, Bi-LSTM with Attention with an AUROC score of 0.959, XGBoost with an AUROC score of 0.935, and Random Forest with an AUROC score of 0.912. From the analysis results, the designed model is able to accommodate interaction among tabular features and the dependencies in time series dataset, thereby allowing for reliable prediction of sepsis disease and proper patient classification. It should be noted, however, that another significant finding from the research was the low number of false negatives produced by the designed model, which is important because failure to diagnose the disease can lead to grave outcomes. The inclusion of missingness indicators in the current model design has not yet been considered, and, therefore, their incorporation would surely improve the effectiveness and efficiency of the designed model, especially because of its clinical importance in sepsis prediction.

References

[1] Rudd, Kristina E., Sarah Charlotte Johnson, Kareha M. Agesa, Katya Anne Shackelford, Derrick Tsoi, Daniel Rhodes Kievlan, Danny V. Colombara et al. "Global, Regional, And National Sepsis Incidence and Mortality, 1990–2017: Analysis for the Global Burden of Disease Study." The Lancet 395, no. 10219 (2020): 200-211.

[2] Singer, Mervyn, Clifford S. Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo et al. "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)." Jama 315, no. 8 (2016): 801-810.

[3] Seymour, Christopher W., Vincent X. Liu, Theodore J. Iwashyna, Frank M. Brunkhorst, Thomas D. Rea, André Scherag, Gordon Rubenfeld et al. "Assessment of Clinical Criteria for Sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)." Jama 315, no. 8 (2016): 762-774.

[4] Kumar, Anand, Daniel Roberts, Kenneth E. Wood, Bruce Light, Joseph E. Parrillo, Satendra Sharma, Robert Suppes et al. "Duration of Hypotension Before Initiation of Effective Antimicrobial Therapy is the Critical Determinant of Survival in Human Septic Shock." Critical care medicine 34, no. 6 (2006): 1589-1596.

[5] Churpek, Matthew M., Frank J. Zadravecz, Christopher Winslow, Michael D. Howell, and Dana P. Edelson. "Incidence and Prognostic Value of the Systemic Inflammatory Response Syndrome and Organ Dysfunctions in Ward Patients." American journal of respiratory and critical care medicine 192, no. 8 (2015): 958-964.

[6] Fleuren, Lucas M., Thomas LT Klausch, Charlotte L. Zwager, Linda J. Schoonmade, Tingjie Guo, Luca F. Roggeveen, Eleonora L. Swart et al. "Machine Learning for the Prediction of Sepsis: A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy." Intensive care medicine 46, no. 3 (2020): 383-400.

[7] Moor, Michael, Bastian Rieck, Max Horn, Catherine R. Jutzeler, and Karsten Borgwardt. "Early Prediction of Sepsis in the ICU Using Machine Learning: A Systematic Review." Frontiers in medicine 8 (2021): 607952.

[8] Reyna, Matthew A., Christopher S. Josef, Russell Jeter, Supreeth P. Shashikumar, M. Brandon Westover, Shamim Nemati, Gari D. Clifford, and Ashish Sharma. "Early

Prediction of Sepsis from Clinical Data: The Physionet/Computing in Cardiology Challenge 2019." Critical care medicine 48, no. 2 (2020): 210-217.

[9] Tonekaboni, Sana, Shalmali Joshi, Melissa D. McCradden, and Anna Goldenberg. "What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use." In Machine learning for healthcare conference, PMLR, 2019, 359-380.

[10] Shashikumar, Supreeth P., Matthew D. Stanley, Ismail Sadiq, Qiao Li, Andre Holder, Gari D. Clifford, and Shamim Nemati. "Early Sepsis Detection in Critical Care Patients Using Multiscale Blood Pressure and Heart Rate Dynamics." Journal of electrocardiology 50, no. 6 (2017): 739-743.

[11] D. Banja, John, Yao Xie, Jeffrey R. Smith, Shaheen Rana, and Andre L. Holder. "Mitigating Bias in Machine Learning Models with Ethics-Based Initiatives: The Case of Sepsis." The American Journal of Bioethics 26, no. 2 (2026): 96-109.

[12] Johnson, Alistair EW, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. "MIMIC-III, A Freely Accessible Critical Care Database." Scientific data 3, no. 1 (2016): 1-9.

[13] Athukorala, V. S., and W. M. K. S. Ilmini. "Explainable AI for Critical Care: A Systematic Review of Interpretable Models for Sepsis and ICU Mortality Prediction." BMC Medical Informatics and Decision Making (2026).

[14] Han, Yupeng, Xiyuan Xie, Jiapeng Qiu, Yijie Tang, Zhiwei Song, Wangyu Li, and Xiaodan Wu. "Early Prediction of Sepsis Associated Encephalopathy in Elderly ICU Patients Using Machine Learning Models: A Retrospective Study Based on the MIMIC-IV Database." Frontiers in Cellular and Infection Microbiology 15 (2025): 1545979.

[15] Kam, Hye Jin, and Ha Young Kim. "Learning Representations for the Early Detection of Sepsis with Deep Neural Networks." Computers in biology and medicine 89 (2017): 248-255.

[16] Lipton, Zachary C., David C. Kale, Charles Elkan, and Randall Wetzel. "Learning to Diagnose with LSTM Recurrent Neural Networks." arXiv preprint arXiv:1511.03677 (2015).

[17] Rajkomar, Alvin, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J. Liu et al. "Scalable and Accurate Deep Learning with Electronic Health Records." NPJ digital medicine 1, no. 1 (2018): 18.

[18] Faust, Oliver, U. Rajendra Acharya, Hojjat Adeli, and Amir Adeli. "Wavelet-based EEG Processing for Computer-Aided Seizure Detection and Epilepsy Diagnosis." Seizure 26 (2015): 56-64.

[19] Chicco, Davide, and Giuseppe Jurman. "The Advantages of the Matthews Correlation Coefficient (MCC) Over F1 Score and Accuracy in Binary Classification Evaluation." BMC genomics 21, no. 1 (2020): 6.

[20] Breiman, Leo. 2001. “Random Forests.” Machine Learning 45 (1): 5–32.

[21] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A Scalable Tree Boosting System." In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, 785-794.

[22] Koller, Daphne, and Nir Friedman. Probabilistic Graphical Models: Principles and Techniques. MIT press, 2009.

[23] Vincent, Pascal, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and Composing Robust Features with Denoising Autoencoders." In Proceedings of the 25th international conference on Machine learning, 2008, 1096-1103.

[24] Komorowski, Matthieu, Leo A. Celi, Omar Badawi, Anthony C. Gordon, and A. Aldo Faisal. "The Artificial Intelligence Clinician Learns Optimal Treatment Strategies for Sepsis in Intensive Care." Nature medicine 24, no. 11 (2018): 1716-1720.

[25] Nemati, Shamim, Andre Holder, Fereshteh Razmi, Matthew D. Stanley, Gari D. Clifffrd, and Timothy G. Buchman. "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU." Critical care medicine 46, no. 4 (2018): 547-553.

[26] Joshi, Yash. “Sepsis Prediction Dataset.”, Accessed March 17, 2026. https://www.kaggle.com/datasets/tea340yashjoshi/sepsis-prediction-dataset.