Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Parallel Deep Learning Framework with Feature-Level Fusion for Early Sepsis Prediction in ICU Settings: Toward Real-Time Clinical Decision-Making

Hadj Ali Elmerahi[1], Baghdad Atmani[1], Fatiha Barigou[1], Belarbi Khemliche[2], Badreddine Errouane[2], Mohammed Bousmaha[2]

[1]*Laboratoire d'Informatique d'Oran (LIO), University of Oran 1 Ahmed Benbella, Oran, Algeria*
[2]*Faculty of medicine, University of Oran 1 Ahmed Benbella, Oran, Algeria*

ARTICLE INFO

Received: 30 Dec 2024

Revised: 05 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Sepsis remains a major cause of ICU mortality, requiring accurate early prediction for clinical decision-making. Conventional deep learning (DL) architectures based on feature-level fusion, even when high-performing, are limited in capturing dynamic feature dependencies and may not ensure real-time applicability. We propose the Parallel Neural Network Fusion Predictor of Sepsis (PNNFPS), an ICU-oriented early prediction framework that employs two Deep Neural Networks (DNN1, DNN2) to process distinct clinical modalities in parallel, with their outputs fused and fed into an artificial neural network (ANN) to estimate sepsis risk. The approach is structured into two complementary phases: (1) offline development and optimization, using 5-fold cross-validation with class-imbalance handling, feature extraction, hyperparameter optimization, automated feature selection, and post-hoc explainability; and (2) online simulation, where the best offline model is deployed in a simulated ICU environment under four alarm policies (low, medium, high, multi-threshold) combined with an 8-hour silencing mechanism to reduce redundant alarms. The resulting alarm streams are evaluated through patient-wise detection, window-wise predictive behavior, alarm burden, alarm reduction, and lead-time analysis. Using the 2019 PhysioNet/Computing in Cardiology Challenge dataset, PNNFPS achieved a competitive offline U-score of 0.673, outperforming previously reported approaches. SHAP-based explainability confirmed clinical consistency, identifying Temperature, Heart Rate, HCO3, pH, WBC, Lactate, and Calcium as key predictive contributors. In online evaluation, the medium-threshold policy provided the most balanced operating point, preserving high patient-wise sensitivity (90.6%) with improved specificity (40.8%) and PPV (18.6%) over the low-threshold policy, an extended lead time of 89 hours before sepsis onset, and substantial alarm reduction (85.64%) via silencing. These findings indicate a favorable balance between patient safety, alarm reliability, and clinician workload, supporting PNNFPS for real-time ICU monitoring and clinical decision support.

Keywords: Early sepsis prediction, DNN, Deep learning, Feature fusion, Alarm policy, Explainability.

INTRODUCTION

Sepsis represents a major public health issue and remains one of the leading causes of death worldwide [1]. In the United States alone, sepsis accounts for nearly 35% of all hospital-related deaths [2]. In addition to its substantial clinical burden, sepsis imposes significant economic costs on healthcare systems, with hospital expenditures estimated at $38.2 billion in 2017, making it the most expensive condition to manage in U.S. hospitals [3]. Despite advances in clinical monitoring and treatment, early prediction of sepsis remains a difficult task, primarily due to

the complex and heterogeneous nature of the clinical data, including vital signs, laboratory measurements and demographic information [4]. In this context, the increasing availability of electronic health records (EHRs) has enabled the application of deep learning (DL) techniques to support early sepsis prediction. These techniques leverage large volumes of heterogeneous data to identify subtle patterns that may precede the clinical onset of sepsis. However, effectively exploiting such data requires appropriate fusion strategies to integrate information from multiple clinical modalities. Accordingly, feature fusion techniques, also referred to as feature combination or integration, play a crucial role in capturing complex interactions among clinical variables by combining complementary clinical information.

In several DL-based prediction systems, early fusion strategies are the most commonly adopted approach, particularly within individual models or sequential hybrid architectures [6], [7]. While such approaches enable unified feature learning at the input level, they obscure the specific contributions of different clinical modalities and still rely partially on hand-engineered feature sets, potentially introducing biases in the learned representations. An alternative strategy is late fusion, which is typically employed in parallel architectures where distinct subsets of features are processed independently by dedicated sub-networks before being integrated at the output level [8]. Despite the promising capabilities of DL models, many existing approaches for early sepsis prediction still rely predominantly on early fusion strategies, which may limit their ability to fully exploit the heterogeneous nature of ICU clinical data and reduce clinical decision explainability, a prevalent 'black box' issue in DL. This lack of transparency poses a critical challenge in healthcare, where clinicians require explainable evidence to trust model predictions and make informed decisions for patient care. Furthermore, a key challenge in early sepsis prediction is the discrepancy between high offline performance and real-time clinical applicability. Models optimized on retrospective data often fail to account for continuous monitoring constraints and alarm management requirements in ICU settings, limiting their practical deployment.

To address these limitations and bridge the gap between theoretical modeling and clinical practice, this study introduces an early prediction framework named Parallel Neural Network Fusion Predictor of Sepsis (PNNFPS). It is a parallel DL–based fusion framework for predicting sepsis six hours before its clinical onset. PNNFPS leverages commonly recorded clinical data, including physiological parameters (e.g., heart rate, blood pressure and temperature) and biochemical data (e.g., lab test results). These data sources are processed through dedicated neural network modules before being integrated through a late fusion mechanism. This framework encompasses three DL sub-networks: two Deep Neural Networks (DNN1 and DNN2) and an Artificial Neural Network (ANN) for prediction. The PNNFPS enables the simultaneous processing of multiple types of historical or Look-Back Windows (LBWs) data [9]. In particularly, the DNN1 operates in parallel with the DNN2 to extract robust feature representations from different clinical data modalities, followed by a feature fusion technique (late fusion) applied to their outputs. Moreover, the ANN is used to generate a final sepsis probability score within a specified timeframe or prediction window (PW) [9].

The proposed approach is structured across two primary phases: (1) Offline Development and Optimization, and (2) Online Simulation Toward Clinical Decision-Making. During the first phase, the PNNFPS framework is constructed through a systematic pipeline prioritized for architectural integrity and predictive performance. This phase incorporates an automated preprocessing module designed to manage missing values and irregular clinical measurements, temporal data transformation to ensure consistent input representation. In particularly, the pipeline integrates feature extraction from 48-hour overlapping ICU windows followed by automated feature selection, resulting in a compact set of clinically relevant features (144 vital-sign and 156 laboratory extracted-features) that preserve both temporal dynamics and physiological variability. In parallel, model optimization is conducted via rigorous hyperparameter tuning combined with 5-fold cross-validation, yielding five PNNFPS models. Additionally, a patient-wise class balancing strategy is applied to address a well-known challenge in ICU datasets, ensuring equitable representation of septic and non-septic cohorts during training. Moreover, post-hoc explainability is integrated within this pipeline to support interpretability of sepsis predictions, while a single decision threshold is optimized on the validation set to convert predicted probabilities into binary clinical decisions.

During the online phase, the highest-performing model identified in offline phase is deployed within a simulated clinical environment that reproduces real-time ICU monitoring by replaying patient data as time-ordered

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

sequential streams. This simulation framework enables the investigation of model behavior under continuous monitoring conditions and supports the generation of clinically actionable alerts. To characterize the clinical utility of PNNFPS from multiple decision-making perspectives, the resulting alarm stream is analyzed under different alerting configurations defined by multiple decision thresholds and an 8-hour silencing policy [10]. Four alarm policies including low, medium, high, and multi-threshold strategies [11] are applied to generate risk-stratified alerts, while the 8-hour silencing mechanism suppresses redundant alarms. The framework is subsequently examined through five complementary alarm-based analyses: patient-wise detection, window-wise predictive behavior, alarm burden, alarm reduction and lead-time performance. Patient-wise analysis reflects classification behavior at the patient-level across entire ICU stays. Window-wise analysis characterizes predictive behavior within clinically relevant pre-sepsis intervals (e.g., [SEPSIS-6h, SEPSIS] and [SEPSIS-48h, SEPSIS]). Alarm burden is quantified as the average daily number of alarms per patient. Alarm reduction measures the decrease in alarm frequency achieved by the silencing strategy. Lead-time performance, defined as the temporal interval between the first model-generated alarm and the clinical onset of sepsis.

The proposed approach was evaluated on the PhysioNet/Computing in Cardiology Challenge 2019 dataset [12] under both offline and online alarm-based settings. In the offline phase, PNNFPS achieved a utility score of 0.673, outperforming previously reported approaches in the literature. SHAP-based explainability further demonstrated the clinical consistency of the framework, identifying Temperature, Heart Rate, HCO3, pH, WBC, Lactate, and Calcium as the most influential predictors of sepsis. In the online phase, PNNFPS demonstrated clinically meaningful performance across all alarm policies. At the patient level, the medium-threshold strategy emerged as the most balanced operating point, achieving a sensitivity of 90.6%, a specificity of 40.8%, and a PPV of 18.6%. Although the high-threshold policy achieved the highest specificity (55.7%), its sensitivity decreased to 82.9%, leaving approximately one-fifth of septic patients undetected. At the window level, the medium-threshold policy maintained sensitivity above 90% across all prediction horizons, while PPV increased from 7.9% at SEPSIS-6h to 47.7% at SEPSIS-48h. Lead-time analysis further showed that alarms were generated, on average, 89 hours before clinical sepsis onset. Under the same policy, the 8-hour silencing mechanism reduced the average alarm burden from 13.79 to 1.98 alarms per patient per day in septic patients, from 7.62 to 1.16 in non-septic patients, and from 8.42 to 1.26 across the overall cohort, corresponding to alarm reductions of 85.64%, 84.78%, and 85.04%, respectively. Collectively, these results indicate a favorable balance between patient detection, alarm reliability, early-warning capability, and clinician workload, supporting the applicability of PNNFPS for real-time ICU monitoring and clinical decision-support systems. The remainder of this paper is organized as follows: Section II reviews the related work in deep learning for sepsis prediction and existing fusion strategies. Section III details the proposed methodology, including the offline development pipeline and the online simulation framework. Section IV presents the experimental results and discusses the clinical implications of the findings and provides concluding remarks regarding future research directions for real-time clinical decision-making.

RELATED WORK

Over the years, numerous definitions have been established to identify sepsis (sepsis onset), including sepsis-1 [13], sepsis-2 [14] and sepsis-3 [4]. While these clinical scoring systems are widely used for screening and diagnosis, they are not designed for predicting early sepsis onset. Nevertheless, many deep learning (DL) models developed for early sepsis prediction often rely on these scoring systems to define sepsis onset, label cases, or construct ground truth during model training.

Deep Neural Network (DNN), a subclass of DL architectures, has been widely employed for early sepsis prediction in several studies, including [6] and [15]. Most of these approaches extract knowledge from ICU data using hierarchical architecture. For example, Erik H Gilbertson [6] developed a DNN-based early warning system. B. T. Lee et al. [34] developed a DNN-based early warning system using graph-based imputation, demonstrating promising clinical applicability but requiring further validation. Shashikumar et al. [16] introduced a feedforward DNN designed to reduce dimensionality and extract salient features from high-dimensional clinical signals. Their model predicts sepsis 4 to 48 hours before clinical suspicion (sepsis-3). It also aims to reduce false alarms by detecting unfamiliar patients or situations arising from erroneous data. The authors reported consistent performance across different levels of care, hospitals and validation cohorts, confirming its applicability in ICU and

emergency department (ED) settings. While these studies have shown that DNNs are effectively identify complex patterns within high dimensional datasets, they struggle to handle temporal dependencies in time series data. This limitation hinders their ability to model the dynamic progression of sepsis, potentially leading to delays in diagnosis and treatment.

Recurrent neural networks (RNNs) address this challenge by learning directly from multivariate time series, minimizing the need for conventional feature extraction. Studies [7], [9] and [17] have demonstrated that RNNs are capable for capturing the time-dependent nature of sepsis progression. For instance, Kam & Kim [7] developed an LSTM model to predict sepsis onset 3 hours before diagnosis (sepsis-2). Their results showed that LSTM outperformed Deep Feed-forward Network (DFN) and InSight [18]. Scherpf et al. [9] introduced a Gated recurrent unit (GRU), which effectively captured temporal dependencies. Nejedly et al. [19] employed an LSTM model and optimized its hyperparameters with a genetic algorithm, demonstrating its robustness in handling sequential ICU data, though performance varied across different test sets. Zhang et al. [20] proposed an explainable LSTM model for predicting sepsis 4 hours prior its onset (sepsis-2) in the emergency department (ED) setting. Despite their ability to handle temporal dependencies, the performance of RNNs remains limited when used individually, as they primarily focus on sequential patterns rather than intricate feature interactions. To overcome this limitation, Sequential or staked hybrid DL models have been explored, combining multiple DL architectures to leverage their complementary strengths. These hybrid DL follow a pipeline approach where the output of one model serves as the input for another, enabling a more structured extraction and transformation of information. Rafiei et al. [21] introduced the Smart Sepsis Predictor (SSP), integrating LSTM and CNN sequentially. Although SSP achieved a high AUROC, its sequential nature in exploring different type of data restricts its ability to effectively account for the unique contributions of each data type.

Parallel hybrid DL models process different data aspects simultaneously through multiple models [8]. Unlike sequential or staked hybrid DL techniques, they do not depend on previous models' outputs. These models use feature fusion technique based on their outputs to form a comprehensive representation (late fusion). Fusion techniques including concatenation, element-wise addition, multiplication and other advanced fusion techniques are detailed in [5]. In our previous work [22], a based parallel LSTM-DNN model was introduced for early prediction of sepsis using ICU datasets. This earlier approach demonstrated reasonable predictive performance and effectively generalized to test data. However, its optimization was limited to tuning the learning rate and the number of neurons per layer while maintaining a default classification threshold of 0.5. This limitation is particularly significant in clinical contexts, where threshold selection directly impacts the balance between sensitivity and specificity and thus the quality of clinical decision-making. Nevertheless, the study highlighted the potential of parallel DL architectures and emphasized the need for more advanced optimization strategies, including threshold optimization, improved feature fusion mechanisms and enhanced model explainability. Wang et al. [23] proposed a multimodal Transformer based architecture for early sepsis prediction, integrating both structured physiological time-series data and unstructured clinical notes. Their framework consists of two parallel modules: the Physiological Time Series Model (PTSM) and the Clinical Notes Model (CNM). The PTSM uses a stack of Transformer encoder layers with multi-head self-attention and feedforward sublayers to model sequential clinical measurements, while the CNM utilizes ClinicalBERT to generate contextualized representations of medical notes. The outputs of both modalities are concatenated and passed through a fully connected neural network, followed by a Softmax layer for binary classification. This parallel and multimodal design enables the model to capture both temporal dynamics and semantic context, demonstrating improved performance in early sepsis prediction compared to unimodal baselines. Other complex hybrid architecture, combining Trees and DL has been introduced by Duan et al. [24] for early sepsis prediction. In particularly, authors developed a Double Fusion Sepsis Predictor (DFSP), which integrates a sequential CNN-GRU network with GBDT models, followed by an ANN model. DFSP employs both early and late fusion techniques [9]. In early fusion, deep features extracted from the sequential CNN-GRU are concatenated with handcrafted features, while in late fusion, prediction from multiple GBDT models are concatenated and refined by an ANN. Notably, each model in the DFSP framework was trained independently as a separate instance, without interdependencies, distinguishing it from parallel models that process data simultaneously. Although, DFSP effectively leverages complementary information, its implementation remains complex and requires careful optimization and validation to ensure robust performance across diverse

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

healthcare settings. This design may limit its suitability for real-time or sequential inference scenarios, where coordinated model interaction and efficient temporal processing are essential.

METHODOLOGY

1- Proposed Approach Overview

This study presents the Parallel Neural Network Fusion Predictor of Sepsis (PNNFPS), a parallel deep learning framework developed to predict sepsis six hours prior to its clinical onset (Sepsis-3), using the 2019 PhysioNet/Computing in Cardiology (CinC) Challenge dataset [12]. An overview of the proposed approach is presented in Figure 1. As shown in this figure, the approach consists of two complementary phases: offline development and online monitoring. In the offline phase (upper panel), vital signs and laboratory measurements from the patient cohort are partitioned into training (80%) and holdout testing (20%) sets, then undergo preprocessing and optimization before being processed through two parallel deep neural networks (DNN1 and DNN2). Their outputs are combined via an addition-based late fusion strategy and fed into an artificial neural network (ANN) to generate sepsis risk scores. Feature contributions are subsequently analyzed using SHapley Additive exPlanations (SHAP) to enhance model transparency and clinical interpretability. The online phase (lower panel) simulates real-time ICU deployment, where patient data are continuously evaluated against a predefined risk threshold (yellow line) to support timely decision-making. For septic patients, the framework is expected to trigger a true alarm within predefined prediction windows (e.g., from Sepsis-48 to onset), enabling early intervention. The figure also illustrates missed alarms, where predicted risk remains below threshold despite subsequent sepsis onset, and a silencing mechanism (muted-speaker icon) that suppresses repetitive alarms within a predefined interval to mitigate alarm fatigue. For non-septic patients, predicted risk generally remains below threshold throughout the ICU stay, though occasional fluctuations may produce false alarms, similarly managed by the silencing strategy. Collectively, the two phases demonstrate how PNNFPS combines accurate early prediction, interpretable decision support, and alarm management to facilitate practical deployment in real-world critical care settings.

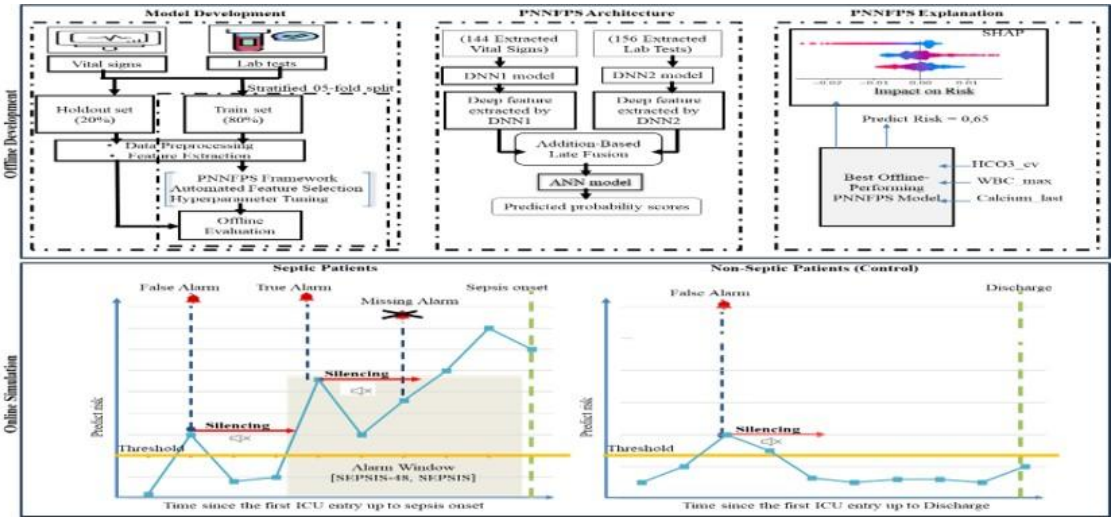

Figure 1. Overview of the proposed approach. **PNNFPS**: Parallel Neural Network Fusion Predictor of Sepsis.

2- Dataset Description

The dataset utilized in this study is the 2019 PhysioNet/Computing in Cardiology (CinC) Challenge dataset [12], publicly available via the PhysioNet platform [25]. It contains de-identified ICU records from 40,336 patients admitted to two institutions (Hospital A: 20,336 patients; Hospital B: 20,000 patients). Clinical data are stored in Pipeline Separated Value (.PSV) files and consist of hourly time-series measurements for 40 variables, including 8 vital signs, 26 laboratory measurements and 6 demographic attributes. The dataset comprises 1,552,210 hourly records. In addition to this publicly available data, the challenge organizers reserved a sequestered test set of 24,819 patients, including additional records from Hospitals A, B and data from an independent Hospital C. Sepsis labels

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

are assigned according to the Sepsis-3 definition [4]. Importantly, the dataset annotates patients as septic up to six hours prior to clinical onset, enabling early prediction modeling. Consistent with real-world ICU data, this dataset exhibits a highly imbalanced. Only 2,932 patients (7.27%), corresponding to 27,916 hourly records were developed sepsis. Moreover, the dataset presents additional challenges, including irregular sampling frequencies, missing values and occasional erroneous. These characteristics informed the design of the preprocessing pipeline described below.

3- Offline Development and Optimization

3.1 Data Preprocessing

Prior to model construction, data preprocessing is essential to ensure high quality inputs. This step is particularly important in sepsis ICU datasets, which are inherently irregular, heterogeneous and frequently incomplete. In this work and following the approach in [26], a systematic preprocessing pipeline is conducted in a fold-aware manner to prevent information leakage. Specifically, stratified 5-fold cross-validation is applied to 80% of the dataset, while the remaining 20% is reserved as an independent holdout set for final test. Patient from all hospitals (A and B) are aggregated into a single Data-frame to facilitate analysis. Timestamps and Patient identifiers are incorporated to preserve the temporal structure of sepsis progression and prevent patient identification overlap. To ensure model robustness, features with prohibitive missing data rates were dropped [27]. Furthermore, administrative and unit-specific variables were excluded to prioritize physiological predictors of sepsis [21]. Consequently, the initial 40-feature set was reduced to 29 retained variables for subsequent analysis. The imputation process is applied to the input data consisting of vital signs and laboratory tests. For each feature, if all values for a given patient are missing, they are replaced with the training fold mean. Any remaining missing values are filled using sequential backward- and forward-fill. To capture the dynamic progression of sepsis, both raw and imputed data were structured into 48-hour overlapping sliding windows with one-hour increments. These windows serve as the basis for feature extraction, as detailed in the following subsection. Furthermore, all numerical features are standardized to ensure consistent scaling across variables.

3.2 Feature Extraction

Previous studies have highlighted the importance of exploiting temporal physiological interactions and missing-data patterns for early sepsis prediction. N.Nesaragi et al. [26] demonstrated that structured temporal representations derived from clinical covariates can effectively model high-dimensional patient dynamics across ICU stays, while B. T. Lee et al. [28] emphasized that missingness patterns, temporal gaps, and inter-variable correlations contain clinically informative signals for critical care prediction tasks. Motivated by these observations, feature extraction in this work was performed on the segmented 48-hour overlapping windows using both raw and imputed data. The objective is to encode complementary aspects of patient state, including data acquisition behavior, temporal physiological dynamics and clinically meaningful interactions (Table 1).

Table 1. Extracted features within the segmented 48-hour overlapping window

Feature Group	Extracted Features	Description
Vital descriptors	**Variables:** HR, MAP, SBP, DBP, O2Sat, Resp, Temp **From raw data:** measurement count, missing ratio, measurement frequency **From imputed data:** last measurement (within 48h window), median, standard deviation, maximum, minimum, coefficient of variation (CV), temporal acceleration (Acc), absolute change (delta), linear trend (slope)	Capture short- and long-term physiological dynamics. Reflect patient instability and temporal fluctuations. Enable detection of progressive deterioration or acute events.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	Multi-scale computation : 4h, 12h, 24h, 48h	
Laboratory descriptors	**Variables:** Glucose, Potassium, Hct, FiO2, Hgb, pH, BUN, WBC, Creatinine, Platelets, Calcium, BaseExcess, Chloride, HCO3, Phosphate, PTT, Lactate, AST, Alkalinephos, Bilirubin direct, SaO2, Bilirubin total. **From raw data:** measurement count, missing ratio, measurement frequency, time since last measurement. **From imputed data:** last measurement, median, standard deviation, maximum, minimum, coefficient of variation (CV), temporal acceleration (Acc), absolute change (delta), linear trend (slope) **Window:** 48h	Characterize organ function and biochemical status. Capture delayed physiological responses and inflammatory processes. Provide complementary information to vital sign dynamics.

Notes: All features are computed within 48-hour overlapping sliding windows with a one-hour stride. Multi-scale features are derived from nested sub-windows (4h, 12h, 24h and 48h) to capture temporal dynamics at different resolutions. Temporal acceleration (Acc) corresponds to the second-order difference, absolute change (delta) represents the difference between the last and first values and slope denotes the linear trend estimated via least-squares regression. Raw-data features reflect data acquisition patterns, while imputed-data features describe physiological evolution

First, features derived from raw data capture measurement patterns, enabling the model to account for missingness structure and sampling irregularities commonly observed in ICU settings. Second, features computed from imputed data encode the temporal evolution of physiological and laboratory variables through statistical summaries and trend-based descriptors. For vital signs, a multi-scale strategy is adopted to capture both short-term fluctuations and long-term trends, while laboratory variables are characterized over the full window to reflect slower biochemical responses.

3.3 Patient-Wise Class Balancing

Several studies have reported that severe class imbalance may bias sepsis prediction models toward the dominant non-septic population, thereby limiting sensitivity to early septic deterioration. B. T. Lee et al. [28] employed selective sampling strategies during training to improve robustness against rapidly evolving septic patterns, whereas Z. Wang *et al.* [29] emphasized the importance of patient-level evaluation protocols and balanced validation settings for reliable clinical deployment. Accordingly, the proposed framework incorporates a stratified, patient-aware under-sampling strategy to prevent model bias toward the majority class. For each positive instance, we selected a corresponding set of negative controls using a hybrid approach:

- Intra-patient Sampling: Negative windows were drawn from the septic patient's own history to capture the transition from a baseline state to sepsis onset.
- Inter-patient Sampling: Complementary negative windows were sampled from a control cohort of non-septic patients to represent general non-septic physiological profiles.

This dual-source sampling ensures the model learns to distinguish sepsis not only from healthy controls but also from a patient's own pre-symptomatic variability. After the sampling process, the dataset was re-indexed and chronologically sorted to ensure that while the class distribution was balanced, the underlying patient-level alignment remained intact for cross-validation and training. Notably, this sampled strategy is used only during the automated feature selection and hyperparameter tuning stages, as we explain below.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.4 Automated Feature Selection and Hyperparameter Optimization

In deep learning–based clinical decision support, selecting the right hyperparameters and features is critical to control model behavior and avoid unnecessary complexity. Traditional approaches such as grid search method test all possible combinations but quickly become inefficient as the search space increases. Random search offers a more practical alternative by exploring the spaces more efficiently with ewer iterations, often discovering good configurations that grid search may miss. Gradient-based can be useful when parameters are continuous, but its applicability remains limited in many clinical settings. Recently, Bayesian optimization [30] has emerged as a powerful solution, as it intelligently guides the search process and significantly reduces the number of required evaluations, making it well-suited for complex models such as those used in early sepsis prediction [27] and [31].

Accordingly, Bayesian optimization is employed to identify the most relevant subset of extracted clinical features and to optimize the configuration of the PNNFPS model. The search is conducted over a high-dimensional space composed exclusively of engineered temporal descriptors derived from vital signs and laboratory variables. The optimization process follows an automated *ask-and-tell* strategy, simultaneously selecting informative features and tuning architectural parameters. This enables the model to retain only the most discriminative extracted features while discarding redundant or less informative ones. In parallel, hyperparameter optimization is conducted over a search space including the number of blocks in each subnetwork (DNN1 and DNN2), batch size, number of units per layer and learning rate. The optimization is performed in a fold-aware manner using 80% of the dataset, ensuring that each fold is treated independently and preventing information leakage between training and validation subsets. The Bayesian optimization process is executed over 400 iterations per fold, where each configuration is evaluated using a single training epoch and assessed based on the Area Under the Precision-Recall Curve (AUPRC) metric on the validation set. The Adam optimizer [32] is employed to minimize the focal loss function [33], enabling robust learning under class imbalance. Distinct hyperparameter configurations are identified for each fold, ensuring systematic and reliable tuning across all data partitions. Once the optimal feature subset and hyperparameters are determined, a second training phase is performed without sampling data. In this stage, the PNNFPS model is retrained for 100 epochs using tow optimized one-dimensional vectors, while fold-specific threshold optimization is also conducted to determine the final clinical decision boundary, ensuring that the model is both statistically optimized and clinically actionable. The first one-dimensional vector represents extracted-features related to vital signs, capturing continuous physiological dynamics, meanwhile the second includes extracted-laboratory features, encoding biochemical and delayed responses.

3.5 Model Architecture

Early sepsis prediction from ICU data involves modeling heterogeneous clinical information, including vital signs and laboratory measurements, which exhibit complex, nonlinear and multi-scale dynamics [28]. A wide range of sequential models have been explored to capture temporal dependencies directly from raw time-series data; however, such approaches often come with increased computational complexity and limited interpretability [21]. For this reason, we adopted two fully connected deep neural networks (DNN1 and DNN2) in a parallel architecture (PNNFPS) to preserve heterogeneity of different input modalities, while enabling the learning of complementary representations and meaningful feature interactions. Figure 2 illustrated the PNNFPS architecture which consists of four components: (1) DNN1 branch, (2) DNN2 branch, (3) an element-wise additive fusion layer and (4) an Artificial Neural Network (ANN). PNNFPS is designed to process simultaneously two vectors of extracted features, corresponding to vital signs and laboratory variables. Each vector is analyzed through a dedicated DNN branch composed of multiple sequential blocks of dense, batch normalization and dropout layers.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

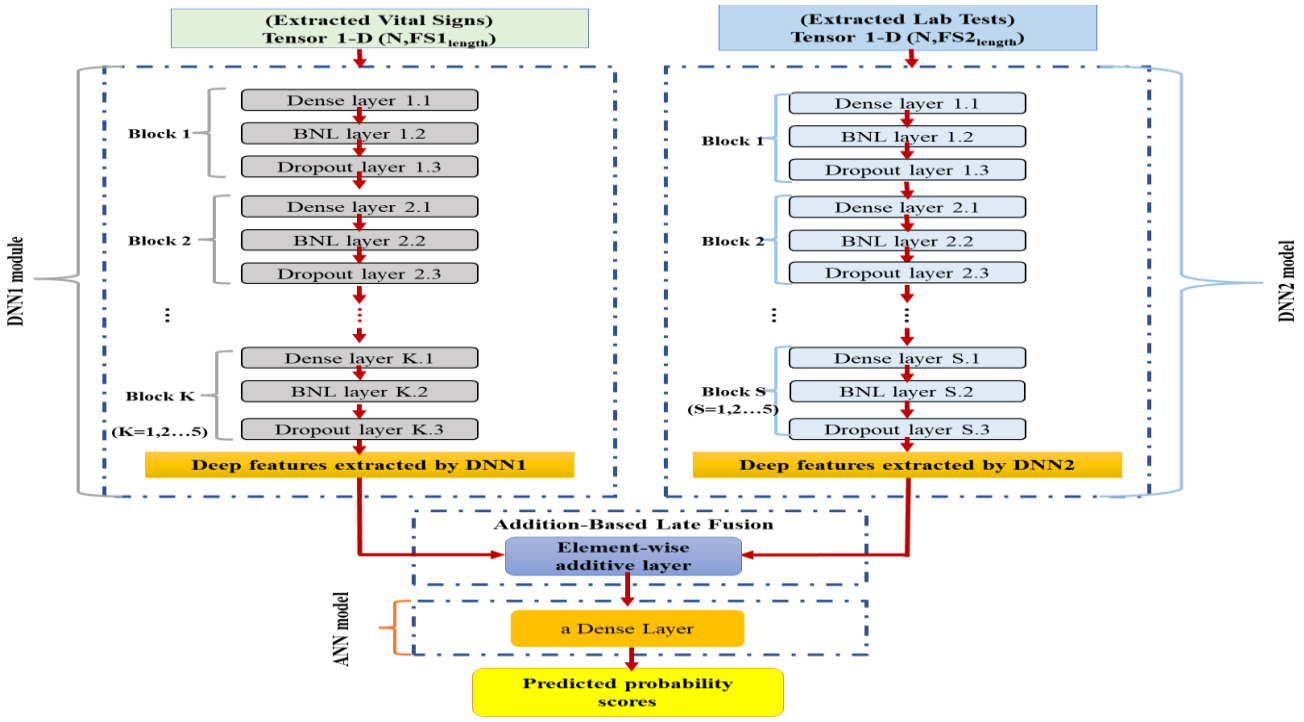

Figure 2. PNNFPS Architecture. **Abbreviations**: DNN1 and DNN2: Two Deep Neural Networks. ANN: an Artificial Neural Network. BNL: Batch Normalization Layer. N: number of samples. FS1 and FS2: two features' sets. $FS1_{length}$ and $FS2_{length}$: number of features in FS1 and FS2 respectively.

DNN1 model: DNN1 processes extracted features derived from vital signs. These features encode temporal dynamics through multi-scale descriptors such as central tendency, variability and trend-related measures computed over the observation window. It consists of several fully connected blocks, each including a dense layer followed by batch normalization and dropout. Batch normalization stabilizes training by maintaining normalized feature distributions, while dropout reduces overfitting and improves generalization. This branch transforms the physiological descriptors into higher-level representations that emphasize clinically relevant patterns associated with sepsis progression.

DNN2 model: The second branch (DNN2) processes extracted features derived from laboratory measurements. DNN2 is composed of different fully connected blocks, each including a dense layer, batch normalization and dropout. This branch focuses on learning nonlinear interactions among laboratory-derived features, enabling the model to capture complex biochemical relationships relevant to sepsis onset. it's worth noting that DNN2 branch operates in parallel with DNN1 branch, enabling the proposed architecture to simultaneously learn representations from multi-scale physiological signals and structured clinical variables. The resulting feature representations are subsequently integrated in the fusion stage of the framework.

Feature Fusion via Element-Wise Additive Layer: As highlighted in Section 2, many DL approaches for early sepsis prediction rely on early fusion strategy, which combines heterogeneous clinical features at the input level [23]. While this approach simplifies model design, it may dilute information when modalities with distinct statistical properties are merged prematurely. Several alternative strategies exist for integrating heterogeneous feature representations at the output level (late fusion), including concatenation-based fusion, element-wise multiplication and attention-based mechanisms [5]. Concatenation preserves modality-specific information but increases feature dimensionality and introduces additional parameters [23]. Attention-based fusion can enhance representation learning by weighting features according to relevance, yet it adds architectural complexity that may reduce transparency and destabilize training—a key concern in clinical decision-support systems [5]. Based on these considerations, the proposed framework adopts element-wise additive fusion [5]. It sums the high-level representations of DNN1 and DNN2 branches. Integrating these modality-specific representations only at the fusion stage allows each branch to extract complementary information while avoiding premature mixing of

heterogeneous features. This design maintains the distinction between multi-scale descriptors and structured clinical features, allowing their respective contributions to be analyzed in a way that supports clinical decision-making processes.

ANN Model: Several prediction mechanisms have been proposed to handle the final output from of the fused feature representation including linear models [18], [34], tree-based model [26] and neural network–based predictors [28]. Linear models are attractive due to their simplicity, interpretability and low computational cost. However, they assume relatively simple relationships between predictors and the outcome. When the inputs correspond to high-level representations learned by deep feature extraction modules, the relationships between these features may involve complex nonlinear interactions that cannot be adequately captured by linear decision boundaries. Tree-based models provide a more flexible alternative, as they naturally represent nonlinear relationships and interactions between variables. However, integrating tree-based classifiers into a DL pipeline remains less straightforward. Unlike neural network layers, which are optimized jointly through backpropagation, tree-based models are typically trained independently [24]. Moreover, combining Tree-based features with deep feature extraction require multi-stage training procedures, additional data transformations and separate model management [24]. In contrast, neural network–based predictors integrate naturally with upstream DL components and allow the entire architecture to be optimized jointly in an end-to-end manner [28]. This property facilitates consistent gradient propagation across all layers of the network and simplifies the training process when deep feature representations are produced by preceding modules. For these reasons, neural network–based predictors are preferred in this work. Specifically, the fused feature representation is provided as input to a final ANN output model, which consists of a fully connected dense layer with a sigmoid activation function. This layer maps the fused feature vector to a scalar probability representing the likelihood 6-hour before of sepsis onset. The sigmoid activation constrains the output to the interval [0, 1], enabling a probabilistic interpretation of the model predictions. Except for the output layer, all hidden layers employ the Rectified Linear Unit (ReLU) activation function. ReLU is widely used in DL architectures due to its computational efficiency and its ability to mitigate the vanishing gradient problem.

3.6 Explainability of the Predictive Model

Deep learning models, although effective for early sepsis prediction, often operate as black boxes, making it difficult for clinicians to explain the rationale behind their predictions. In ICU settings, explainability is therefore crucial to support clinical decision-making and identify key features contributing to sepsis risk. Several post-hoc explainability methods have been proposed for neural networks. Techniques such as LIME [35], DeepLIFT [36] and Integrated Gradients [37] are widely used to estimate feature contributions; however, they may exhibit variability or rely on local approximations. In contrast, SHAP (SHapley Additive exPlanations) [38] provides a more consistent and theoretically principled framework, offering both global and local explanations based on Shapley values derived from cooperative game theory, thereby ensuring greater interpretability and reliability .

Due to its solid theoretical foundation and its ability to provide consistent feature attributions, SHAP is employed in this study to analyze the behavior of the best offline-performing PNNFPS model. Specifically, SHAP is implemented via the Gradient Explainer, which approximates Shapley values for differentiable models using expected gradients [38]. A balanced subset of septic and non-septic patients is selected as the background distribution for SHAP estimation. This choice is consistent with prior work emphasizing that explanations should be interpreted relative to a representative reference distribution to ensure clinical relevance [38]. Model interpretability is investigated from both local and global perspectives. At the local level, SHAP values assign an importance score to each feature for individual hourly predictions, enabling the identification of patient-specific risk factors contributing to predicted sepsis probabilities. At the global level, feature importance is assessed by averaging the absolute SHAP values across all samples, providing an overall ranking of the most influential clinical variables. To ensure robust and unbiased interpretation, both local and global analyses are conducted on a randomly selected, balanced subset of both septic and non-septic patients from the test set. SHAP values are computed at multiple levels of the proposed architecture, including the DNN1 branch to capture the contribution of vital sign descriptors, the DNN2 branch to assess laboratory-derived features and the fusion layer to evaluate the relative importance of each feature group in the final prediction.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2- Online Simulation Toward Clinical Decision-Making

2.1 Simulation of Real-Time Monitoring

In ICU settings, early sepsis prediction systems based on deep learning (DL) require timely and reliable evaluation to support clinical decision-making. In traditional offline evaluation, models are assessed on the entire test data under a controlled condition. While this evaluation is commonly used to identify generalization performance prior deployment, it may overestimate model performance and obscure practical constraints related to real-world temporal data availability. Based on these considerations, the best offline-performing model is deployed within an online simulation framework. In this setting, raw test data are processed sequentially where predictions are produced at each timestamp as patient data become available, reflecting real-time monitoring. For septic patients, only data from the time of first ICU entry up to the clinical onset of sepsis are considered. For non-septic patients, data from the entire ICU length of stay (ICULOS) is used. This simulation follows the methodology in [22], where both septic and non-septic patients are fully included, unlike [9], which restricts evaluation to selected look-back windows. In addition, to support real-time clinical decision-making, the generated predictions are assessed continuously using a multi-threshold scheme in conjunction with a silencing alarm policy, as detailed in the following subsection.

2.2 Alarms and Continuous Predictions

In a real-time monitoring setting, continuous hourly predictions generate a stream of risk scores rather than a single decision point. Using a single threshold to trigger these alarms, as in offline binary classification, can lead to excessive false alerts due to temporal variability and noise, ultimately contributing to alarm fatigue in clinical practice [16]. For these reasons, multiple clinically-informed thresholds were employed in this work, combined with silencing policy alarm to enable risk-stratified alarm generation in a continuous monitoring setting. This policy means that the model would be silenced for a period of time after an alarm was triggered [10]. The silencing period was set to 8 h following recent research [11]. To this end, a data-driven calibration strategy was adopted based on the distribution of model outputs on the validation set to calculate these thresholds. Specifically, all predicted hourly probabilities produced by the model were aggregated across patients and time points, excluding any post-sepsis periods. Thresholds were then determined by targeting predefined alarm rates per patient, expressed in terms of expected alarms per day under a no-silencing scenario [11]. Following established clinical simulation protocols, three operating points were defined: a low threshold corresponding to approximately three alarms per patient per day, a medium threshold corresponding to one alarm per patient per day and a high threshold corresponding to one alarm per patient per week [11]. These alarm frequencies were converted into percentile-based cutoffs over the empirical distribution of predicted probabilities. This approach ensures that the thresholds are adaptive to the model's calibration and output distribution, while maintaining clinically interpretable alarm frequencies independent of specific patient trajectories.

Based on these thresholds and the silencing policy, alarm system of PNNFPS is analyzed under four alarm policies: low-, medium-, high- and multi-threshold policy [11]. The low-, medium-, and high-threshold strategies are single-threshold approaches, where each decision threshold is applied independently. In these cases, an alarm is triggered when the risk score exceeds the selected threshold, and all subsequent alarms are suppressed during the 8-hour refractory (silence) period. As a low-threshold alarm example, an alarm is triggered when the patient's risk score exceeds the low-threshold at time (t=10h). After this trigger, an 8-hour silence period is activated, during which no additional alarms are generated even if the risk score at time (t=14h) continues to increase. In contrast, the multi-threshold strategy introduces a hierarchical mechanism where higher risk levels can override the silence constraint, allowing alarm escalation when patient condition worsens. This ensures that critical clinical deterioration is not masked, while still maintaining control over alarm redundancy. For example, an initial alarm may occur at time (t=10h) when the patient's risk score exceeds a low-threshold. Subsequently, at time (t=14h), the risk score increases further and surpasses a medium-threshold, which triggers an escalated alarm even though the system is still within the original 8-hour silence window.

All the four policies are assessed at two levels including patient-wise and window-wise. At the patient-wise level, patients would be identified as septic cases if at least one prediction during their ICU stays and whose probability

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

was higher than a single threshold and non-septic cases if not. At the window-wise level, clinical relevance is further evaluated by considering different early-warning horizons, reflecting varying clinician requirements for sepsis anticipation [39]. Specifically, we set four different expected alarm windows including [SEPSIS-6, SEPSIS], [SEPSIS-24, SEPSIS], [SEPSIS-36, SEPSIS] and [SEPSIS-48, SEPSIS]. We defined an expected window as the time period in which alarms should be triggered preceding the sepsis onset. Any alarm within this window was considered a true positive (TP). Alarms occurring outside of that window were treated as false positives (FP), while the absence of alarms within the window was considered as a false negative (FN). To assess the operational behavior of the alarm system under continuous monitoring, alarm statistics were computed separately for three patient groups separately: all patients, septic patients and non-septic patients. These statistics specifically include the average daily number of alarms per patient and the corresponding reduction rate achieved through the silencing strategy. In addition, the system's early warning capability was evaluated by analyzing the temporal interval between the first generated alarm and the clinical sepsis onset. Detailed definitions of these evaluation metrics are provided in the following subsection.

2.3 Model Evaluation Metrics

The proposed PNNFPS framework performance is evaluated under both offline and online settings. In the offline setting, PNNFPS is assessed using the custom-defined Utility score (U score) [12], which is a clinically oriented metric that jointly evaluates both prediction accuracy and timeliness. Specifically, U score rewards early correct sepsis predictions while penalizing late detections, false alarms, and missed cases, thereby reflecting the clinical value of early intervention. In the online simulation setting, generated alerts are analyzed under different decision thresholds while applying an 8-hour silencing policy. Evaluation is conducted at two complementary levels: window-wise and patient-wise. At the window-wise level, predictive behavior is assessed across multiple prediction horizons using sensitivity (proportion of true sepsis windows correctly detected) and positive predictive value (PPV; proportion of alarms correctly associated with sepsis). At the patient-wise level, classification outcomes derived from the generated alerts are used to compute sensitivity (proportion of septic patients correctly identified), specificity (proportion of non-septic patients correctly excluded) and PPV (proportion of alerted patients who are truly septic). Furthermore, the average number of alarms per day per patient is computed for septic patients, non-septic patients and the overall population. To quantify the effectiveness of the alarm management strategy, alarm reduction was calculated by comparing the average number of alarms per patient per day before and after application of the 8-hour silencing policy. The reduction was reported as the percentage decrease in alarm burden achieved through suppression of redundant alerts. Finally, the system's early warning capability is calculated in term of the lead time from septic patients, defined as the average interval between the first alarm and the clinical sepsis onset.

RESULTS AND DISCUSSION

1- Offline & Online Evaluation Cohorts

This subsection describes the final cohorts used for offline and online evaluation, extracted from the PhysioNet dataset. For both evaluations, the cohorts represent the fold-data (Fold 1) that yield to the best performing model. The offline cohort includes all imputed patient trajectories (2932 septic vs 37404 non-septic) that are used for training model. In contrast, the online cohort corresponds to the raw hold-out test fold (245 septic vs 1644 non-septic) which we used in the simulated real-time monitoring with the best offline-trained model. To ensure a strictly predictive evaluation and maintain clinical relevance, the online cohort is subject to two critical constraints: first, only patients with an ICU length of stay (LOS) greater than 48 hours are selected to allow for sufficient baseline monitoring. Second, all data points following the clinical sepsis onset are excluded for septic patients.

Within the balanced offline cohort, the automated feature selection based on Bayesian optimization yielded to 144 features derived from 4 original vital signs (HR, MAP, O2Sat, Temp), 156 features extracted from 12 laboratory parameters (Glucose, Calcium, Chloride, HCO3, WBC, Hgb, Hct, Platelets, Creatinine, pH, Lactate, PTT). Vital signs contribute 36 features per variable, combining acquisition information, instantaneous state and multi-scale temporal descriptors. Acquisition includes 3 features (measurement count, missing ratio, measurement frequency), while the instantaneous state is represented by 1 feature (last measurement). Temporal descriptors are computed

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

over four windows (4h, 12h, 24h, 48h), each providing 8 features (mean, std, min, max, slope, delta, cv, acc). Laboratory variables contribute 13 features per parameter, composed of acquisition descriptors, statistical summaries and temporal dynamics: 4 acquisition features (measurement count, missing ratio, measurement frequency, time since last measurement), 5 statistical features (last, mean, std, min, max) and 4 temporal features (slope, delta, cv, acc). Table 2 summarizes the baseline characteristics of both cohorts with only the most influential clinical variables corresponding to the extracted features. These features were identified through SHAP-based feature importance analysis and include the top 10 predictors from vital signs and top 10 predictors from laboratory measurements. Septic patients show distinct statistical patterns compared to non-septic patients across these variables, reflecting their clinical deterioration profiles. Locking to the offline cohort, reduced variability in some characteristics like pH and Lactate among non-septic patients is attributed to imputation toward central tendency values, leading to lower dispersion.

Table 2. Baseline Characteristics cohorts for both offline and online settings

Setting	Characteristic	Septic	Non-Septic	p value
Offline	Age (years)	64 (51.58–74.11)	63 (51.00–73.64)	0.105
	Temp (°C)	37.00 (36.59–37.47)	36.81 (36.50–37.15)	<0.001
	HR (bpm)	88.28 (77.40–99.22)	82.60 (73.08–92.43)	<0.001
	Platelets (10^3/μL)	179.00 (132.07–241.89)	184.50 (146.06–244.49)	<0.001
	HCO3 (mmol/L)	23.47 (21.37–25.52)	23.47 (22.93–24.83)	<0.001
	Calcium (mg/dL)	8.36 (7.90–8.65)	8.40 (8.06–8.75)	<0.001
	WBC (10^3/μL)	10.86 (8.80–14.30)	10.20 (7.70–12.80)	<0.001
	Lactate (mmol/L)	0.54 (0.54–0.54)	0.54 (0.54–0.54)	<0.001
	pH	7.39 (7.37–7.42)	7.39 (7.39–7.39)	<0.001
	Gender (Male)	1739 (59.3%)	20827 (55.7%)	<0.001
	ICU Type (Medical)	2282 (77.8%)	25602 (68.4%)	<0.001
Online	Age (years)	65 (52.01–73.00)	65 (53.00–75.00)	0.288
	Temp (°C)	37.03 (36.65–37.51)	36.84 (36.53–37.20)	<0.001
	HR (bpm)	86.18 (78.47–96.40)	83.72 (73.90–93.92)	<0.001
	Platelets (10^3/μL)	187.17 (118.57–239.33)	188.50 (135.75–250.00)	0.091
	HCO3 (mmol/L)	23.40 (21.12–25.90)	23.75 (21.44–26.05)	0.369
	Calcium (mg/dL)	8.22 (7.90–8.66)	8.30 (7.90–8.75)	0.202
	WBC (10^3/μL)	11.70 (9.49–14.56)	10.20 (7.70–13.30)	<0.001
	Lactate (mmol/L)	0.19 (0.14–0.28)	0.19 (0.15–0.27)	0.631
	pH	7.41 (7.37–7.44)	7.39 (7.35–7.42)	<0.001
	Gender (Male)	142 (58.0%)	895 (54.4%)	0.335
	ICU Type (Medical)	181 (73.9%)	1151 (70.0%)	0.245

a: Baseline characteristics are reported at the patient level after aggregating repeated hourly measurements using patient-wise grouping. Continuous variables are summarized as median (interquartile range) and categorical variables as n (%). p-values were computed using the Mann–Whitney U test for continuous variables and the chi-

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

square test for categorical variables. Statistical significance was defined at $p < 0.001$.

2- Offline Performance

Table 3 presents the offline evaluation results, where fold-specific metric indicates that Fold 1 achieved the highest U score (0. 673) demonstrating superior predictive performance compared to other folds. However, Fold3 exhibited comparable performance with U score of 0. 672, showing only a marginal difference. Across the five folds, PNNFPS achieved an average U-score of 0.650 ± 0.024, with Fold 1 selected as the leader model for subsequent online evaluation. These results indicate that the proposed PNNFPS framework generalizes effectively to unseen data within a controlled evaluation setting.

Table 3. Offline evaluation performance of each test fold.

Fold N°	**1 (leader model)**	**3**	**4**	**2**	**5**	**Mean ± Std**
U score	0.673	0.672	0.659	0.636	0.612	0.650 ± 0.024

3- Comparative Offline Analysis with Existing Approaches

Table 4 summarizes a comparative analysis of PNNFPS against existing approaches under the offline setting, highlighting their key findings and predictive performance. The comparison includes studies such as [40] and [41], which correspond to submissions to the 2019 PhysioNet/CinC Challenge dataset as well as more recent studies including [22], [42] and [43]. The results demonstrate that PNNFPS achieved an U score of 0.672, representing 56% improvement over the highest performing literature benchmark (Morrill et al [42], U=0.434). These improvements reflect the combined impact of Bayesian optimization of both hyperparameter and the automated feature selection, adaptive threshold tuning and the parallel DL architecture introduced in this study.

Table 4. Comparative offline Analysis of the Proposed PNNFPS Framework and Existing Solutions

Reference	**Methodology / Model**	**U-Score**
Morrill et al (2020) [42]	A signature-based transform model	0.434
Li et al. (2020) [43]	Time-phased model (CNN-LSTM)	0.430
Chang et al. (2019) [40]	Engineered features from missing data (TCN)	0.417
Vicar et al. (2019) [41]	Feature normalization and missing value handling (Deep LSTM)	0.350
Our Previous Work (2024) [22]	Basic parallel LSTM-DNN	0.137
Proposed Approach	**PNNFPS Framework (Parallel Deep Learning)**	**0.673**

4- Explainability Analysis

Figure 3 present both local and global SHAP analysis of the best performing PNNFPS model, based on the balanced test-fold data, with the corresponding sampled training-fold serving as the background set. This figure is organized as Beeswarm plot with two panels: Panel (a) illustrates the temporal dynamics of vital signs, Panel (b) shows the contribution of laboratory features. Both panels display the 10 most influential extracted-features for each branch (DNN1 and DNN2). By looking at the vertical ranking, we gain a global view of what the model values most across the entire ICU population, while the individual dots provide a local window into how the model treats a specific patient at a specific hour. The color coding acts as a clinical signal: red represents high measurements and blue represents low ones. When these colors shift to the right, they indicate a positive impact that pushes the risk score toward a sepsis diagnosis. For example, seeing a cluster of red dots on the right for HCO3_CV confirms that high levels are a major risk driver, while blue dots on the right for Calcium would indicate that low value is being flagged as a critical warning sign. The long-tails on certain features are particularly insightful, as they highlight variables that may not be important for everyone but can become life-saving predictors for a small subset of critically ill

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

patients. Referring to panel (a), Temperature (Temp) and Heart Rate (HR) emerge as the main constitutors to early sepsis detection. Notably, the model's sensitivity is concentrated during the earliest time steps (hours 4–24), indicating that the PNNFPS architecture effectively captures early physiological alterations preceding sepsis onset. This temporal behavior is consistent with clinical evidence showing that early sepsis is frequently associated with early compensatory physiological responses, particularly thermoregulatory and cardiovascular alterations such as fever and tachycardia, which may precede significant laboratory abnormalities [44]. In summary, this contribution pattern highlights the interpretability of the parallel architecture. Unlike single-representation approaches [20], PNNFPS quantifies physiological and biochemical contributions separately, with DNN1 accounting for 42% and DNN2 for 57% of the overall predictive contribution, thereby providing more transparent risk attribution.

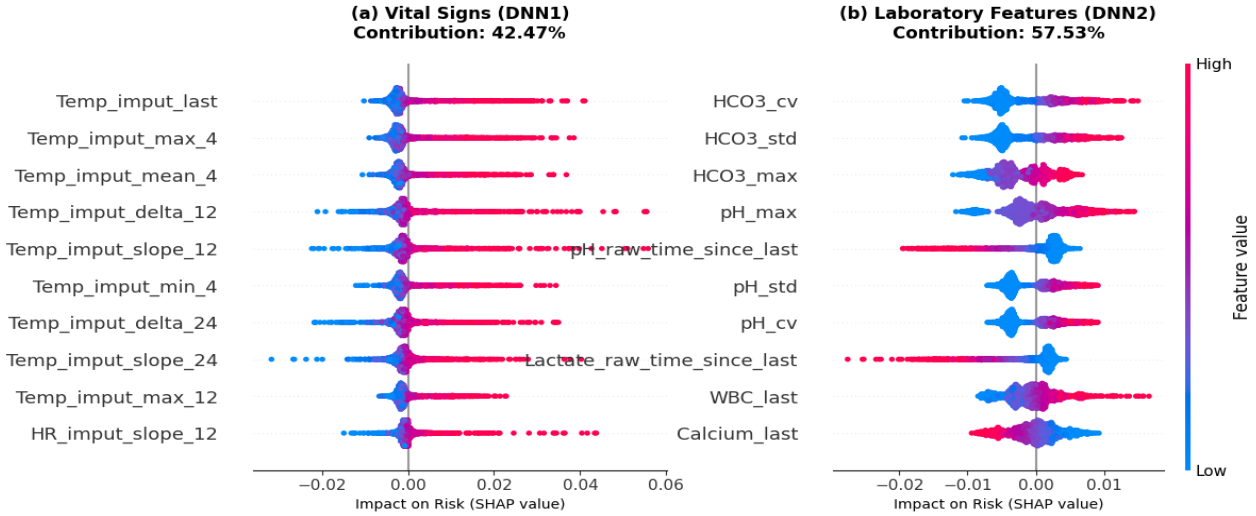

Figure 3. Multimodal Explainability of Sepsis Risk: A Dual-Branch Global and Local Feature Analysis. Features containing the pattern 'imput' are derived from imputed data, whereas the others originate from raw measurements.

5- Online Alarm-Based Analysis

Table 5 presents the alarm-level performance of PNNFPS in a simulated real-time ICU setting, focusing on patient-wise detection across complete ICU stays and window-wise predictive behavior across multiple prediction horizons. All results are reported under the 8-hour alarm silencing policy, which suppresses redundant alarms during continuous monitoring.

From the window-wise perspective, predictive performance improved as the observation horizon expanded from SEPSIS-6h to SEPSIS-48h. For example, under the medium-threshold policy, sensitivity remained consistently high across all horizons, while PPV increased markedly from 7.9% at SEPSIS-6h to 47.7% at SEPSIS-48h. Similar trends were observed for the other alarm policies. These results indicate that a larger proportion of septic patients becomes detectable as the prediction window extends, leading to higher sensitivity across all alarm policies. At the same time, the progressive increase in PPV demonstrates that alarms generated closer to clinical onset are more reliable and clinically meaningful. This behavior is consistent with the progressive nature of sepsis, whereby physiological and biochemical abnormalities become increasingly pronounced as onset approaches, allowing the model to identify high-risk states with greater confidence. Conversely, prediction at earlier stages remains inherently more uncertain because the underlying signs of deterioration are often subtle and less specific [45]. When comparing evaluation metrics across alarm policies, a clear trade-off was observed between sensitivity and PPV. The high-threshold policy consistently achieved the highest PPV, increasing from 8.0% at SEPSIS-6h to 48.1% at SEPSIS-48h, indicating superior alarm reliability. However, this improvement was accompanied by a reduction in sensitivity, which ranged from 63.7% at SEPSIS-6h to 94.2% at SEPSIS-48h, resulting in a greater proportion of missed septic patients. In contrast, the low- and multi-threshold policies maximized patient detection, achieving

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

the highest sensitivities at SEPSIS-48h (97.8% and 97.9%, respectively), but with slightly lower PPV values (46.3% and 46.0%, respectively), reflecting a higher false-alarm burden. The medium-threshold policy provided the most balanced compromise between early identification of septic patients and reduction of false alarms. It maintained consistently high sensitivity across all prediction horizons (>90%) while achieving PPV values close to those obtained with the high-threshold policy, particularly at longer horizons (25.3%, 37.0%, and 47.7% at SEPSIS-24h, SEPSIS-36h, and SEPSIS-48h, respectively). This finding is consistent with previous studies showing that effective early warning systems require a balance between sensitivity and alarm burden [46], [47], [48].

Table 5. Patient-wise detection and window-wise predictive performance of PNNFPS across multiple prediction horizons under the 8-hour alarm silencing policy in a simulated real-time ICU setting.

Evaluation perspective	Alarm Policy	Window	Sensitivity	PPV
Window-wise (Septic patients)	Low	SEPSIS-6h	71.4%	7.4%
		SEPSIS-24h	94.5%	24.0%
		SEPSIS-36h	96.8%	35.7%
		SEPSIS-48h	97.8%	46.3%
	Medium	SEPSIS-6h	97.8%	7.9%
		SEPSIS-24h	91.3%	25.3%
		SEPSIS-36h	94.9%	37.0%
		SEPSIS-48h	96.4%	47.7%
	High	SEPSIS-6h	63.7%	8.0%
		SEPSIS-24h	88.5%	26.0%
		SEPSIS-36h	92.1%	37.4%
		SEPSIS-48h	94.2%	48.1%
	Multi-threshold	SEPSIS-6h	73.1%	7.3%
		SEPSIS-24h	94.6%	23.7%
		SEPSIS-36h	97.0%	34.9%
		SEPSIS-48h	97.9%	46.0%
		Specificity	**Sensitivity**	**PPV**
Patient-wise (All patients)	Low & multi-threshold	18.7%	96.3%	15.0%
	Medium	40.8%	90.6%	18.6%
	High	55.7%	82.9%	21.8%

At the patient level, PNNFPS demonstrated robust detection capability, with at least one clinically relevant alarm generated for most septic patients. As the decision threshold increased, sensitivity gradually decreased from 96.3% under the low-threshold policy to 90.6% and 82.9% under the medium- and high-threshold policies, respectively. Conversely, specificity improved substantially from 18.7% to 40.8% and 55.7%, while PPV increased from 15.0% to 18.6% and 21.8%, respectively. The multi-threshold policy produced identical patient-level results to the low-threshold policy because patient classification depended solely on the occurrence of at least one alarm during the ICU stay; therefore, additional escalated alarms did not alter the final classification outcome. These results further support the findings obtained from the window-wise analysis. While the low-threshold policy maximized septic patient detection (96.3% sensitivity), it was associated with lower classification reliability (18.7% specificity and

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

15.0% PPV). Conversely, the high-threshold policy improved specificity (55.7%) and PPV (21.8%) but reduced sensitivity to 82.9%, leaving approximately one-fifth of septic patients undetected. Consistent with the trends observed across prediction horizons, the medium-threshold policy provided the most balanced patient-level performance, maintaining high sensitivity (90.6%) while substantially improving specificity (40.8%) and PPV (18.6%). These findings reinforce the selection of the medium-threshold policy as the reference operating point for continuous ICU monitoring.

Table 6 presents the complementary alarm-level performance of PNNFPS, focusing on alarm burden (average daily number of alarms per patient), alarm reduction achieved through the silencing strategy, and lead-time performance reflecting the timeliness of sepsis detection. Alarm burden is reported both with and without the 8-hour alarm silencing policy, in order to quantify its effect on reducing redundant alarms during continuous monitoring.

Table 6. Alarm burden, alarm reduction, and lead-time performance of PNNFPS with and without the 8-hour alarm silencing policy in a simulated real-time ICU setting

Alarm Policy (Lead time)	Septic patient Without → with (Reduction)	Non-Septic patient Without → with (Reduction)	All-patients Without → with (Reduction)
Low (99h)	15.30 → 2.24 (85.36%)	9.53 → 1.53 (83.94%)	10.28 → 1.62 (84.24%)
Medium (89h)	13.79 → 1.98 (85.64%)	7.62 → 1.16 (84.78%)	8.42 → 1.26 (85.04%)
High (88h)	12.72 → 1.80 (85.85%)	6.74 → 0.97 (84.15%)	7.52 → 1.08 (85.64%)
Multi-threshold (99h)	15.31 → 2.39 (84.39%)	9.54 → 1.68 (82.39%)	10.28 → 1.77 (82.78%)

Across all alarm policies, the 8-hour silencing mechanism substantially reduced alarm burden while preserving early warning capability. For septic patients, the average number of alarms per patient per day decreased from 12.72–15.31 to 1.80–2.39, corresponding to an alarm reduction of approximately 84–86%. Similar reductions were observed in non-septic patients (82–84%) and across the overall cohort (82–85%). These findings indicate that a large proportion of alarms generated during continuous monitoring were redundant and could be safely suppressed without compromising the delivery of clinically relevant alerts. Such reductions are particularly important in ICU environments, where excessive alarm exposure is a major contributor to alarm fatigue and reduced responsiveness to critical alarms [49]. Lead-time analysis further demonstrated that PNNFPS provides clinically meaningful early warning of sepsis. Across all alarm policies, alarms were generated approximately 88–99 hours before clinical onset. The low-threshold policy achieved the longest mean lead time (99 hours), whereas the medium- and high-threshold policies maintained comparable lead times of 89 and 88 hours, respectively. This extended warning horizon is clinically relevant because early recognition of sepsis enables timely intervention and has been associated with improved patient outcomes [50], [51].

Within this context, the medium-threshold policy remained the most balanced operating point across all alarm-based analyses. In addition to providing favorable patient-wise and window-wise performance, it maintained a substantial lead time of 89 hours while reducing the average alarm burden from 13.79 to 1.98 alarms per patient per day in septic patients (85.64% reduction), from 7.62 to 1.16 in non-septic patients (84.78% reduction), and from 8.42 to 1.26 across the overall cohort (85.04% reduction). These results indicate that most redundant alarms can be eliminated without sacrificing the model's ability to provide early warning. This consistency across multiple evaluation perspectives strengthens the robustness of the selected operating point and supports its applicability for real-time ICU monitoring and clinical decision-support systems.

CONCLUSION

This study presented PNNFPS, a parallel deep learning framework designed for early sepsis prediction in ICU settings, integrating heterogeneous clinical information, including vital signs and laboratory measurements through a late feature-level fusion strategy. Interpretability was ensured through SHAP-based analysis, which identified the most influential clinical variables contributing to each prediction, thereby enhancing transparency

and supporting clinician trust in model-driven decision making. This is further reinforced by the parallel architecture of PNNFPS, where two dedicated branches independently model physiological (DNN1, 42%) and biochemical (DNN2, 57%) features, enabling a clearer attribution of the risk source behind each prediction.

Evaluation on the 2019 PhysioNet/Computing in Cardiology Challenge dataset demonstrated consistent improvements under offline settings, with PNNFPS achieving a utility score (U-score) of 0.673, corresponding to a 56% improvement over the best-performing benchmark reported in the literature. To better reflect practical ICU deployment conditions, four alarm policies (low, medium, high, and multi-threshold) were incorporated alongside an 8-hour silencing mechanism to reduce redundant alarms and mitigate alarm fatigue under continuous monitoring. Among all policies, the medium-threshold policy emerged as the most balanced operating point, reducing the average daily alarm burden from 13.79 to 1.98 (85.64% reduction) in septic patients, from 7.62 to 1.16 (84.78% reduction) in non-septic patients, and from 8.42 to 1.26 (85.04% reduction) across the overall cohort, while maintaining an extended lead time of 89 hours before sepsis onset. Although PNNFPS was originally trained to predict sepsis 6 hours before clinical onset, clinically relevant alarms were observed substantially earlier, indicating that the model's learned representations generalize beyond its nominal training horizon and offer a wider window for clinical intervention.

Despite these promising results, several limitations must be acknowledged. First, the study relies exclusively on the PhysioNet/CinC 2019 dataset, which may limit generalizability across different patient populations and clinical settings. Second, the online evaluation was conducted in a simulated rather than a live ICU environment, meaning that real-world deployment factors — such as data latency, missing values in streaming data, and clinician interaction patterns — were not fully captured. Third, the class imbalance inherent to sepsis prediction datasets, despite mitigation strategies applied during training, may still influence model behavior in real-world conditions where sepsis prevalence varies. Finally, prospective clinical validation remains necessary to confirm the robustness and practical impact of the framework. Future work will therefore focus on prospective validation across multi-center hospital datasets to assess generalizability, and on integrating PNNFPS into hospital clinical decision support systems to evaluate its practical feasibility and clinical impact on patient management and ICU outcomes.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENT

This work is part of the Medical Decision-Making Support Based on Artificial Intelligence and Reasoning for Intensive Care (AIR4AIC) project, conducted by the Artificial Intelligence and Reasoning (AIR) team in collaboration with the reanimation service of the University Hospital Establishment (UHE) of Oran, Algeria.

REFRENCES

[1] N. Ocampo-Quintero, P. Vidal-Cortés, L. del Río Carbajo, F. Fdez Riverola, M. Reboiro-Jato, and D. Glez-Peña, “Enhancing sepsis management through machine learning techniques: A review,” *Medicina Intensiva*, vol. 46, no. 3, pp. 140–156, 2022, doi: 10.1016/j.medin.2020.04.003.

[2] C. Rhee *et al.*, “Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014,” *JAMA*, vol. 318, no. 13, p. 1241, 2017, doi: 10.1001/jama.2017.13836.

[3] L. Liang, B. Moore, and A. Soni, “National inpatient hospital costs: The most expensive conditions by payer.” 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK561141/

[4] M. Singer *et al.*, “The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3),” *JAMA*, vol. 315, no. 8, p. 801, 2016, doi: 10.1001/jama.2016.0287.

[5] N. Mungoli, “Adaptive feature fusion: Enhancing generalization in deep learning models.” 2023. doi: 10.48550/arXiv.2304.03290.

[6] E. Gilbertson, K. Jones, A. Stroh, and B. Whitaker, "Early detection of sepsis using feature selection, feature extraction, and neural network classification," in *Computing in Cardiology Conference (CinC)*, 2019. doi: 10.22489/CinC.2019.386.

[7] H. J. Kam and H. Y. Kim, "Learning representations for the early detection of sepsis with deep neural networks," *Computers in Biology and Medicine*, vol. 89, pp. 248–255, 2017, doi: 10.1016/j.compbiomed.2017.08.015.

[8] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Computer Vision and Image Understanding*, vol. 156, pp. 117–127, 2017, doi: 10.1016/j.cviu.2016.10.010.

[9] M. Scherpf, F. Gräßer, H. Malberg, and S. Zaunseder, "Predicting sepsis with a recurrent neural network using the MIMIC III database," *Computers in Biology and Medicine*, vol. 113, p. 103395, 2019, doi: 10.1016/j.compbiomed.2019.103395.

[10] Z. Wang and B. Yao, "Multi-branching temporal convolutional network for sepsis prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 876–887, 2022, doi: 10.1109/JBHI.2021.3092835.

[11] M. (A. M.) Van Den Berg *et al.*, "Development and clinical impact assessment of a machine-learning model for early prediction of late-onset sepsis," *Computers in Biology and Medicine*, vol. 163, p. 107156, Sep. 2023, doi: 10.1016/j.compbiomed.2023.107156.

[12] M. Reyna, A. Josef, A. J. Jeter, B. Shashikumar, J. Westover, and M. Nemati, "Early prediction of sepsis from clinical data: The PhysioNet/Computing in Cardiology Challenge 2019," *PhysioNet*, 2019, doi: 10.13026/V64V-D857.

[13] R. C. Bone *et al.*, "Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis," *Chest*, vol. 101, no. 6, pp. 1644–1655, 1992, doi: 10.1378/chest.101.6.1644.

[14] M. M. Levy *et al.*, "2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference," *Critical Care Medicine*, vol. 31, no. 4, pp. 1250–1256, 2003, doi: 10.1097/01.CCM.0000050454.01978.3B.

[15] Z. Wang and B. Yao, "Multi-branching temporal convolutional network for sepsis prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 876–887, 2022, doi: 10.1109/JBHI.2021.3092835.

[16] S. P. Shashikumar, G. Wardi, A. Malhotra, and S. Nemati, "Artificial intelligence sepsis prediction algorithm learns to say 'I don't know,'" *npj Digital Medicine*, vol. 4, no. 1, p. 134, 2021, doi: 10.1038/s41746-021-00504-6.

[17] D. Zhang, C. Yin, K. M. Hunold, X. Jiang, J. M. Caterino, and P. Zhang, "An interpretable deep-learning model for early prediction of sepsis in the emergency department," *Patterns*, vol. 2, no. 2, p. 100196, 2021, doi: 10.1016/j.patter.2020.100196.

[18] T. Desautels *et al.*, "Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach," *JMIR Medical Informatics*, vol. 4, no. 3, p. e28, 2016, doi: 10.2196/medinform.5909.

[19] P. Nejedly, F. Plesinger, I. Viscor, J. Halamek, and P. Jurak, "Prediction of sepsis using LSTM with hyperparameter optimization with a genetic algorithm," in *Computing in Cardiology Conference*, 2019. doi: 10.22489/CinC.2019.022.

[20] D. Zhang, C. Yin, K. M. Hunold, X. Jiang, J. M. Caterino, and P. Zhang, "An interpretable deep-learning model for early prediction of sepsis in the emergency department," *Patterns*, vol. 2, no. 2, p. 100196, 2021, doi: 10.1016/j.patter.2020.100196.

[21] A. Rafiei, A. Rezaee, F. Hajati, S. Gheisari, and M. Golzan, "SSP: Early Prediction of Sepsis Using Fully Connected LSTM-CNN Model," *Computers in Biology and Medicine*, vol. 128, p. 104110, 2021, doi: 10.1016/j.compbiomed.2020.104110.

[22] H. A. Elmerahi, B. Atmani, F. Barigou, B. Khemliche, B. Errouane, and M. Bousmaha, "Parallel LSTM-DNN Fusion Model for Early Prediction of Sepsis in Intensive Care Units," in *4th International Conference on Embedded & Distributed Systems (EDiS)*, IEEE, 2024, pp. 43–48. doi: 10.1109/EDiS63605.2024.10783411.

[23] Y. Wang, Y. Zhao, R. Callcut, and L. Petzold, "Integrating physiological time series and clinical notes with transformer for early prediction of sepsis." 2022. doi: 10.48550/arXiv.2203.14469.

[24] Y. Duan *et al.*, "Early prediction of sepsis using double fusion of deep features and handcrafted features," *Applied Intelligence*, vol. 53, pp. 17903–17919, 2023, doi: 10.1007/s10489-022-04425-z.

[25] "PhysioNet, Challenge 2019 (version 1.0.0), https://physionet.org/ challenge/2019/#challenge-data, accessed: Sep. 07, 2021 (2019)."

[26] N. Nesaragi, S. Patidar, and V. Aggarwal, "Tensor learning of pointwise mutual information from EHR data for early prediction of sepsis," *Computers in Biology and Medicine*, vol. 134, p. 104430, 2021, doi: 10.1016/j.compbiomed.2021.104430.

[27] M. Fu, J. Yuan, M. Lu, P. Hong, and M. Zeng, "An ensemble machine learning model for the early detection of sepsis from clinical data," in *Computing in Cardiology Conference*, 2019. doi: 10.22489/CinC.2019.317.

[28] B. T. Lee, O.-Y. Kwon, H. Park, K.-J. Cho, J.-M. Kwon, and Y. Lee, "Graph convolutional networks-based noisy data imputation in electronic health record," *Critical Care Medicine*, vol. 48, no. 11, pp. e1106–e1111, 2020, doi: 10.1097/CCM.0000000000004583.

[29] Z. Wang *et al.*, "A methodological systematic review of validation and performance of sepsis real-time prediction models," *NPJ Digit Med*, vol. 8, no. 1, p. 190, Apr. 2025, doi: 10.1038/s41746-025-01587-1.

[30] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian Optimization of Machine Learning Algorithms," 2012, *arXiv*. doi: 10.48550/ARXIV.1206.2944.

[31] F. Chollet, "Keras." 2023. [Online]. Available: https://github.com/fchollet/keras

[32] B. Lee, K. Cho, O. Kwon, and Y. Lee, "Improving the performance of a neural network for early prediction of sepsis," in *Computing in Cardiology Conference*, 2019. doi: 10.22489/CinC.2019.162.

[33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 2999–3007. doi: 10.1109/ICCV.2017.324.

[34] Q. Mao, D. Jay, K. Hoffman, M. Calvert, and D. E. Das, "Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU," *BMJ Open*, vol. 8, no. 1, p. e017833, 2018, doi: 10.1136/bmjopen-2017-017833.

[35] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," Aug. 09, 2016, *arXiv*: arXiv:1602.04938. doi: 10.48550/arXiv.1602.04938.

[36] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning Important Features Through Propagating Activation Differences," Oct. 12, 2019, *arXiv*: arXiv:1704.02685. doi: 10.48550/arXiv.1704.02685.

[37] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic Attribution for Deep Networks," Jun. 13, 2017, *arXiv*: arXiv:1703.01365. doi: 10.48550/arXiv.1703.01365.

[38] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," 2017, *arXiv*. doi: 10.48550/ARXIV.1705.07874.

[39] L. Liu, H. Wu, Z. Wang, Z. Liu, and M. Zhang, "Early prediction of sepsis from clinical data via heterogeneous event aggregation," in *Computing in Cardiology Conference*, 2019. doi: 10.22489/CinC.2019.157.

[40] Y. Chang *et al.*, "A Multi-Task Imputation and Classification Neural Architecture for Early Prediction of Sepsis from Multivariate Clinical Time Series," presented at the 2019 Computing in Cardiology Conference, Dec. 2019. doi: 10.22489/CinC.2019.110.

[41] T. Vicar, J. Hejc, P. Novotna, M. Ronzhina, and R. Smisek, "Sepsis Detection in Sparse Clinical Data Using Long Short-Term Memory Network with Dice Loss," presented at the 2019 Computing in Cardiology Conference, Dec. 2019. doi: 10.22489/CinC.2019.341.

[42] J. Morrill, A. Kormilitzin, A. Nevado-Holgado, S. Swaminathan, S. Howison, and T. Lyons, "Utilization of the signature method to identify the early onset of sepsis from multivariate physiological time series in critical care monitoring," *Critical Care Medicine*, vol. 48, no. 4, pp. e976–e981, 2020, doi: 10.1097/CCM.0000000000004510.

[43] X. Li *et al.*, "A Time-Phased Machine Learning Model for Real-Time Prediction of Sepsis in Critical Care," *Critical Care Medicine*, vol. 48, no. 10, pp. e884–e888, Oct. 2020, doi: 10.1097/CCM.0000000000004494.

[44] M. Cecconi, L. Evans, M. Levy, and A. Rhodes, "Sepsis and septic shock," *The Lancet*, vol. 392, no. 10141, pp. 75–87, Jul. 2018, doi: 10.1016/S0140-6736(18)30696-2.

[45] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman, "An interpretable machine learning model for accurate prediction of sepsis in the ICU," *Critical Care Medicine*, vol. 46, no. 4, pp. 547–553, 2018, doi: 10.1097/CCM.0000000000002936.

[46] M. P. Sendak *et al.*, "Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study," *JMIR Med Inform*, vol. 8, no. 7, p. e15182, Jul. 2020, doi: 10.2196/15182.

[47] K. E. Henry, D. N. Hager, P. J. Pronovost, and S. Saria, "A targeted real-time early warning score (TREWScore) for septic shock," *Sci. Transl. Med.*, vol. 7, no. 299, Aug. 2015, doi: 10.1126/scitranslmed.aab3719.

[48] W. L. Chua, K. D. B. Rusli, and L. M. Aitken, "Early warning scores for sepsis identification and prediction of in-hospital mortality in adults with sepsis: A systematic review and meta-analysis," *Journal of Clinical Nursing*, vol. 33, no. 6, pp. 2005–2018, Jun. 2024, doi: 10.1111/jocn.17061.

[49] A. Sowan, "Effective dealing with alarm fatigue in the intensive care unit," *Intensive and Critical Care Nursing*, vol. 80, p. 103559, Feb. 2024, doi: 10.1016/j.iccn.2023.103559.

[50] A. Kumar *et al.*, "Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock*:," *Critical Care Medicine*, vol. 34, no. 6, pp. 1589–1596, Jun. 2006, doi: 10.1097/01.CCM.0000217961.75225.E9.

[51] L. Evans *et al.*, "Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021," *Critical Care Medicine*, vol. 49, no. 11, pp. e1063–e1143, Nov. 2021, doi: 10.1097/CCM.0000000000005337.