IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 December 2023, accepted 27 December 2023, date of publication 2 January 2024,
date of current version 9 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3349242

RESEARCH ARTICLE

Advancing Early Detection of Sepsis With Temporal Convolutional Networks Using ECG Signals

MERVE APALAK, (Member, IEEE), AND KAMRAN KIASALEH, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA

Corresponding author: Merve Apalak (Merve.Apalak@utdallas.edu)

This work was supported in part by the Computational Resources provided by the BioHPC Supercomputing Facility located at the Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas.

ABSTRACT In the hours leading up to the onset time of sepsis, the autonomic nervous system presents sub-clinical indicators of disease that may not be observable to providers. The objective of this research is to create an interpretable sepsis prediction algorithm utilizing continuous electrocardiography (ECG) signals, with the aim of implementing it in patient monitoring systems for individuals in intensive care units (ICU). We develop an early sepsis detection algorithm utilizing two datasets; in particular, the Medical Information Mart for Intensive Care (MIMIC-III) Clinical Dataset and MIMIC-Waveform Database. We carry out a systematic approach to selecting ECG segments of superior quality that are recorded in highly dynamic intensive care unit environments. Later, we use the single-lead ECG waveform to investigate the potential of heart rate variability (HRV) for continuous monitoring. In this study, 715 patients were included, of whom 65 are sepsis patients labeled with recent sepsis definition on an hourly basis. The predictive potential of the critical features is visualized to assist the interpretation of the model in a clinical practice. Moreover, since we are framing the early sepsis prediction as a supervised time series classification task, we evaluate the model performance by implementing Temporal Convolutional Networks (TCN). Performance analysis reported with varying prediction windows preceding sepsis onset time using area under the receiver-operating-characteristic curve (AUROC) and area under the precision-recall curve (AUPRC). The deep learning model delivers promising results by leveraging time series with the use of temporal convolutions. Our findings reveal that the HRV characteristics of adults can be a valuable indicator for continuous sepsis monitoring in an ICU. Finally, this research work adds to the field of early sepsis detection by providing an annotated continuous waveform dataset from the MIMIC-Waveform Database, which is made accessible to the public.

INDEX TERMS Sepsis, temporal convolutional networks, ECG signal, heart rate variability, continuous monitoring, clinical decision support.

I. INTRODUCTION

Sepsis is a life-threatening medical emergency that is often caused by bacterial, viral, or fungal infections but can also result from other sources of infection or inflammation. It can rapidly lead to tissue damage, organ failure, and death if not recognized and treated in a timely matter [1]. The increase in sepsis incidence constitutes a growing public health concern. Despite recent promising medical advances, sepsis is still recognized as a leading cause of death in hospitals [2].

Recent studies have found that early intervention and recognition of sepsis can significantly reduce the high mortality and morbidity rates. This is addressed in recent clinical and observational studies, which have revealed that the risk of mortality rises with each hour of delay in administering antibiotics [3]. However, indiscriminate and prolonged use of antibiotics is the leading factor for the

The associate editor coordinating the review of this manuscript and approving it for publication was Vincenzo Conti.

proliferation of antibiotic-resistant infections, a concern that the World Health Organization (WHO) has identified as one of the top ten global public health threats [4]. Hence, the importance of timely recognition and initiation of treatment has a fundamental role not only in reducing sepsis-related mortality and morbidity, but also in preventing the adverse impact of antibiotic usage.

Recently, the use of machine learning methods in this space has gained momentum through the use of health data [5]. It can naturally tackle the immense and complex electronic health records by learning the correlation of the patterns in data which in turn is applied for early sepsis prediction [6], [7], [8]. For instance, authors in [6] built a prediction model based on Weilbull-Cox proportional hazards model employing clinical data. In their study, Shashikumar et al. [7]) showed that features extracted from multiscale time series of heart rate (HR) and mean arterial blood pressure (MAP) achieved an Area Under the Curve (AUC) of 0.80 in forecasting the onset of sepsis by four hours. Furthermore, in [9], authors validated combinations of six vital sign measurements and their temporal changes using a well-established sepsis detection algorithm called InSight. More recently, authors investigated the PhysioNet/Computing in Cardiology Challenge 2019 data [10] for early sepsis prediction [11]. As is typical in clinical databases, this dataset displays a significant lack of information, as up to 80% of the data contains missing values. Consequently, considerable effort is dedicated to addressing the challenges associated with handling missing data in order to attain high prediction performances [12], [13], [14], [15]. As a result, identifying sepsis in its early stages with clinical and laboratory data remains a challenging problem due to their natural constraints.

Given the challenges associated with clinical data, researchers have redirected their focus towards alternative data sources, with the electrocardiogram (ECG) emerging as a prominent choice [16]. In [17], the authors created and trained multiple classifiers using extracted features from continuous ECG and arterial blood pressure (ABP) signals. The objective was to recognize the pathological condition of sepsis by utilizing features extracted from physiological waveform within the intensive care unit (ICU) environment. Another study revealed a reduction in various heart rate variability (HRV) parameters among septic patients who did not survive [18]. More recently, the study presented predictive monitoring of late-onset sepsis (LOS) in infant patients by characterizing the prognostic potential of features obtained from ECG recordings such as HRV, the respiratory waveform, and lethargy scores [19]. Despite the limited size of their dataset, their findings suggest that HRV features exhibit promising potential for the early detection of sepsis. Leon et al. [20] introduced a technique that incorporates HRV features along with visibility graph indices into the feature set. Their study demonstrated that the inclusion of these physiological signal characteristics resulted in improved predictive accuracy for LOS in neonates.

Convolutional Neural Networks (CNNs) have made substantial contributions to the healthcare field. Several studies have demonstrated the effectiveness of CNNs, particularly analyzing physiological signals, yielding notably accurate and consequential results [16], [21], [22]. In [23], the Medical Information Mart for Intensive Care (MIMIC)-III dataset was used for the proposed method where the Gaussian process adapter framework was combined with temporal convolutional networks (TCNs) to evaluate predictive performance of sepsis. For instance, Lombardi et al. [24] trained a well-known CNN-based model called ResNet [25] utilizing raw fingertip photoplethysmography (PPG) time-series from MIMIC-III database. While discussing the importance of physiological signals for continuous monitoring of sepsis patients, this study is also investigating selection strategies for dataset preparation based on the utilization of PPG data. Recently, authors developed a method using 12 leads ECG signals to demonstrate the outcomes of applying ECG to predict infection in patients [26].

Similar to the efforts mentioned earlier, our research focuses on predicting early sepsis. However, in contrast to existing methods, our primary objective is to assess the potential of ECG signals in introducing a novel technique for predicting sepsis in the adult population. To our best knowledge, this is the first study working with temporal convolution networks and ECG signals for early sepsis detection. The primary contributions of this work are as follows.

- We propose a systematic workflow for selecting high quality ECG segments applicable for training an early sepsis detection system built on single-lead ECG signals. The assumption is to remove the low-quality ECG signals, thereby improving the performance of predictive model through the elimination of outliers in the waveform database;
- We extract the traditional HRV features to predict sepsis in selected population. We report the feature attributes to gain insights into which features change and impact leading up to the sepsis onset;
- We present a sequence to sequence modeling by implementing TCNs. The deep learning model realises the temporal correlation between features owing to its effective memory without any recursive layer. Additionally, it results in superior performance;
- Finally, we provide an annotated ECG database for the early detection of sepsis facilitating research in the community focused on sepsis detection.

II. MATERIALS AND METHODS

A. DATASET DESCRIPTION

This work uses the MIMIC-III database, which is a publicly available critical care database that consists of electronic health records (EHRs) of over 40,000 patients who had been admitted to critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. The database contains a diverse information such as demographics, diagnoses,

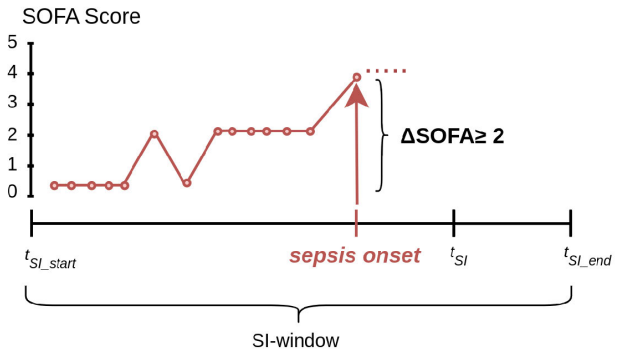

FIGURE 1. Illustration of Sepsis-3 criteria. First, suspicion of infection (SI) time is calculated. According to [1], if antibiotic was administered before the culture test, then the test must be ordered within 24 h. Conversely, if the culture test is obtained first, then the antibiotic treatment has to be started within 72 h. The window of -48h to 24h around first SI is used for checking 2 points of deterioration in SOFA score. The time when the condition is met, it is decided as the onset time of sepsis.

procedures, medications, laboratory tests, vital signs, and other clinical data. It is publicly accessible on PhysioNet [27]. The MIMIC-III Waveform Database Matched Subset is a companion to the MIMIC-III clinical database, which carries various physiological signals such as ECG, ABP, PPG, and respiration. It has been continuously recorded in real-time from bedside monitors and time-aligned with the MIMIC-III clinical database. The waveform records of 10,282 ICU patients are matched to the MIMIC-III clinical database, making their corresponding clinical records also accessible.

B. SEPSIS LABEL DEFINITION

In the context of sepsis diagnosis, there are standard clinical scores to assess the severity of sepsis, namely SIRS, NEWS, or MEWS [28], [29], [30]. However, these scores are not for continuously evaluating the risk of developing sepsis. Recently, Sepsis-3 criteria have been announced and updated to the definition and diagnostic criteria for sepsis [1]. The criteria determine the onset time of sepsis further; it requires suspicion of infection time and Sequential Organ Failure Score (SOFA). We follow the SI definition to calculate the suspicion of infection time suggested in [31]. Sepsis-3 criteria recommend that sepsis onset time be defined as when an increase of at least 2 points in the SOFA score has occurred in the suspicion of infection window. Figure 1 illustrates our Sepsis-3 application. We slightly revise the clinical database queries shared by [23] and [27] on an hourly basis in order to define the sepsis onset time.

C. DATASET PREPROCESSING

It is important to mention that this work is accommodating both the clinical database and the waveform database. Each database is used for a specific task such as annotation and training. The clinical database is utilized to determine the onset time of sepsis by implementing Sepsis-3 criteria. Afterwards, the available ECG recordings are annotated in accordance with the calculated onset time. The main use of the clinical database is to extract demographics, sepsis onset time, and to cross-check for ICU admission and discharge time. No features such as vitals, lab measurements, or nurse/doctor notes are used as inputs to the prediction model. The predictive model is solely trained with annotated waveform database. The proposed preprocessing approach consists of three main phases depending on the accessibility and quality of the ECG recordings. Figure 2 depicts the detail of the preprocessing.

1) PATIENT FILTERING

In our settings, we refine our cohort with the following manner. First, patients who are under the age of 15 are discarded. Secondly, patients with missing ICU admission or discharge s are excluded. Additionally, we abandon the patients listed with the CareVue system due to a lack of lab measurements since we need lab measurements to calculate the suspicion of the infection time. After filtering the data, the sepsis onset time is calculated. ICD-9 is a coding system to track diagnoses and procedures associated with hospital utilization which is also provided for every patient in the clinical database. Finally, we look into ICD-9 codes to further guarantee that the control patients are not sepsis patients before ICU admission. The final population in clinical database comprises 1725 sepsis patients and 17790 control patients. This query setting allows us to annotate the waveform database and identify the patient population.

We encounter many challenges during our analysis on waveform database. The data set contains 10,282 distinct ICU patients and each recording reported with a header file. The header files are necessary to access the recordings. However, we find that 187 patients do not have a header file. The ECG recordings have a variable length, ranging from hours to days. Since the waveform database is a subset of the clinical database, only 447 sepsis patients' waveforms are available in the waveform database. The sepsis onset times and the code were made publicly available https://figshare.com/s/ee778918d7b31c653fe2. Additionally, we count 5319 patients who did not experience sepsis. Frequently, there are multiple waveform record pairs associated with a given patient's record. The challenge is that even though these record pairs belong to a particular ICU stay, they have a long time gap between them. Therefore, if there are multiple recordings for a certain ICU stay, we only consider the first recording or the ones recorded closet to sepsis onset time. It is found that the database contains at least one ECG waveform for every patient but available ECG leads show differences. Consequently, we restrict our cohort to patients whose ECG waveform was measured by lead II. Furthermore, some patients have been admitted to the ICU on more than one occasion; therefore, we only consider one admission for these patients. We also compare the ICU *in-time* and *out-time* with record start and end times. It was discovered

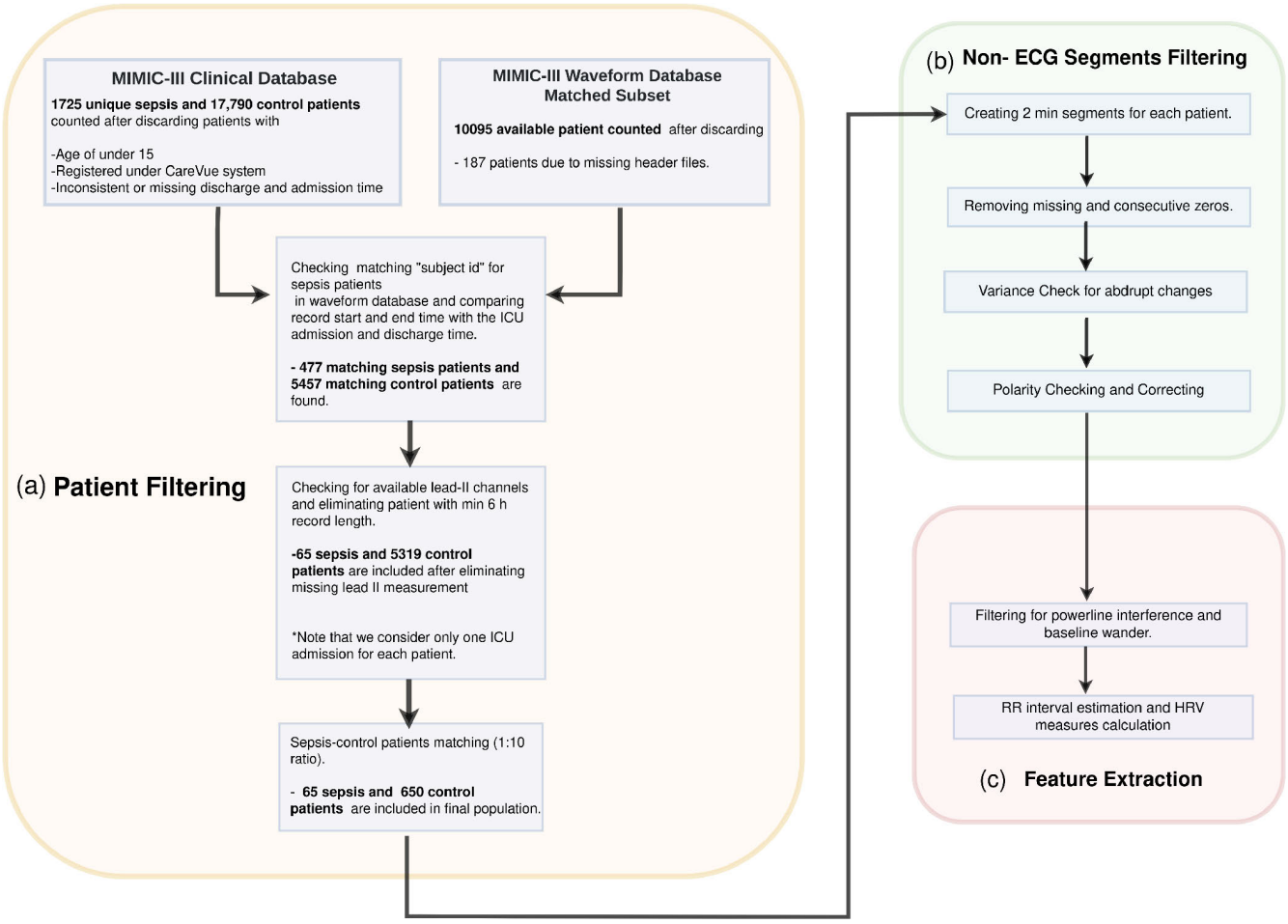

FIGURE 2. Block diagram of the signal processing and feature extraction. (a) An illustration of patient filtering. It starts finding sepsis and control patients in the MIMIC-III clinical database, then finds the matching patients in the MIMIC-III Waveform database. Next, it selects patients involving ECG lead-II recordings while searching for the minimum 6h recordings. It ends with the case-control sampling. (b) In block b, the elimination of non-ECG segments is given. (c) Feature extraction flow starts with noise filtering and continues with HRV measurement extraction.

that even though *subject_id* is detected in both datasets, there is a mismatch due to multiple ICU admissions. For instance, *subject_id=188* was admitted eight times to the ICUs (MICU and SICU). The patient underwent sepsis at the time when his admission time was recorded as "2161-07-01 19:44:00," but records for this stay are not available in the waveform database. The available ECG record belongs to the admission time was "2161-12-09-17-50."

At each hour of an ICU stay, our prediction model focuses to predict whether a sepsis is going to happen during the next -6h to 0h. Figure 3 illustrates the prediction windows settings and label assignment. We assigned the label 0 for each hour of controls. Including 12 hours after onset time can potentially increase the chances of learning the significant features of the sepsis. For the early prediction cases, we include the data from beginning of the recording up to 12 hours after sepsis onset. We follow the same approach as in [10] for labeling. Labels are 1 if $t \geq t_{onset} - t_{pred}$ and 0 if $t < t_{onset} - t_{pred}$. t_{pred} stands for the prediction windows. Moreover, we eliminate the patients developing sepsis earlier than 6 hours. In order to reach similar maximal length of the ICU stay, we extract

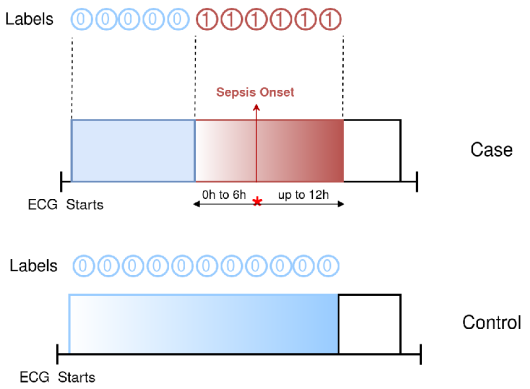

FIGURE 3. Illustration of label composition for single sepsis and control patient. For sepsis patients, we consider the early detection task of 0h to 6h preceding sepsis onset. For example, if the prediction task is 6h before sepsis onset, we label time steps 6 hours before up until 12 hours after sepsis onset as 1. The window is limited to a minimum of 6h.

up to 48 hours of recordings from the beginning of the ECG recordings for the control patients. The overall sepsis prevalence is roughly 10% in the clinical dataset. Hence,

TABLE 1. Description of HRV features.

HRV Features	Description
MeanNN	The mean of the RR intervals.
SDNN	The standard deviation of the RR intervals.
RMSSD	The root mean square of the differences between successive R-R intervals.
SDSD	The standard deviation of the successive differences between RR intervals.
MinNN	The minimum of the RR intervals. [32]
MaxNN	The maximum of the RR intervals. [32]
pNN50	Percentage of successive RR intervals that differ by more than 50 ms.
Prc80NN	The 80th percentile of the RR intervals. [33]
TINN	Baseline width of the RR interval histogram.
LF	The spectral power of low frequencies.
LFn	The normalized low frequency.
HF	The spectral power of high frequencies.
HFn	The normalized high frequency.
SD1	Poincaré plot standard deviation perpendicular the line of identity.
SD2	Poincaré plot standard deviation along the line of identity.
SD1/SD2	SD1/SD2 ratio.
S	Area of the ellipse.
SampEn, ApEN Entropy	Sample and approximate entropy measure the regularity and complexity of a time series.

we apply the 1:10 ratio in our study of cohorts to maintain realistic prediction problem. Table 2 summarizes the statistics for the final population.

2) FILTERING NON-ECG SEGMENTS

Noise detection and quality assessment of ECG signals have been intensively studied in [34] and [35]. Identification of poor-quality ECG signals is essential, particularly in a highly chaotic environment, such as ICUs, given that noisy ECG signals may lead to false alarms and inaccurate diagnosis. We point out that ECG recordings are impaired by a number of issues, such as sudden signal disappearance, a high number of missing values, bad electrode contact, and others. We implement a pipeline to identify the ECG segments from highly corrupted noisy segments.

- We start by dividing each recording into 2-minute segments. Next, if there are any missing values (indicated by NaN) in the target segment, we discard that segment;
- Furthermore, segments with consecutive zero values more than 400 ms are discarded;
- Additionally, some segments in the ECG waveform suddenly disappear. These segments shows a low variance for the 2-minute segments;
- Next, we realize that there are many inverted ECG signals in the waveform database. We, hence, detect them, then we correct the polarity of inverted ECG signals.

Figure 4 depicts an example of the abrupt signal change, which shows low variance for certain duration. After considering all, there are some records still showing non-ECG waveforms. For these records, we perform a manual inspection. In order to keep a track of time steps, we monitor the non-ECG segment filtering for every 1 hour. If we are required to remove all the 2-minute segments in the 1 hour window, we fill corresponding time step with missing values to impute in the further steps.

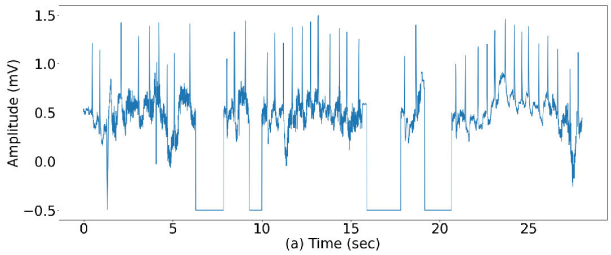

FIGURE 4. 30 s sample recording illustrates the abrupt signal change for Subject ID= 69339.

3) FEATURE EXTRACTION

First, we filter the data to remove typical artifacts due to environmental or biological sources, such as power line interference and baseline wander. We use a notch filter at 50 Hz cutoff and a high-pass Butterworth filter at 0.5 Hz. Before calculating the HRV measurements, QRS complexes are detected based on the steepness of the gradient of the waveform. Once the QRS complex is located, the R-peaks can be identified as local maxima within the QRS complex.

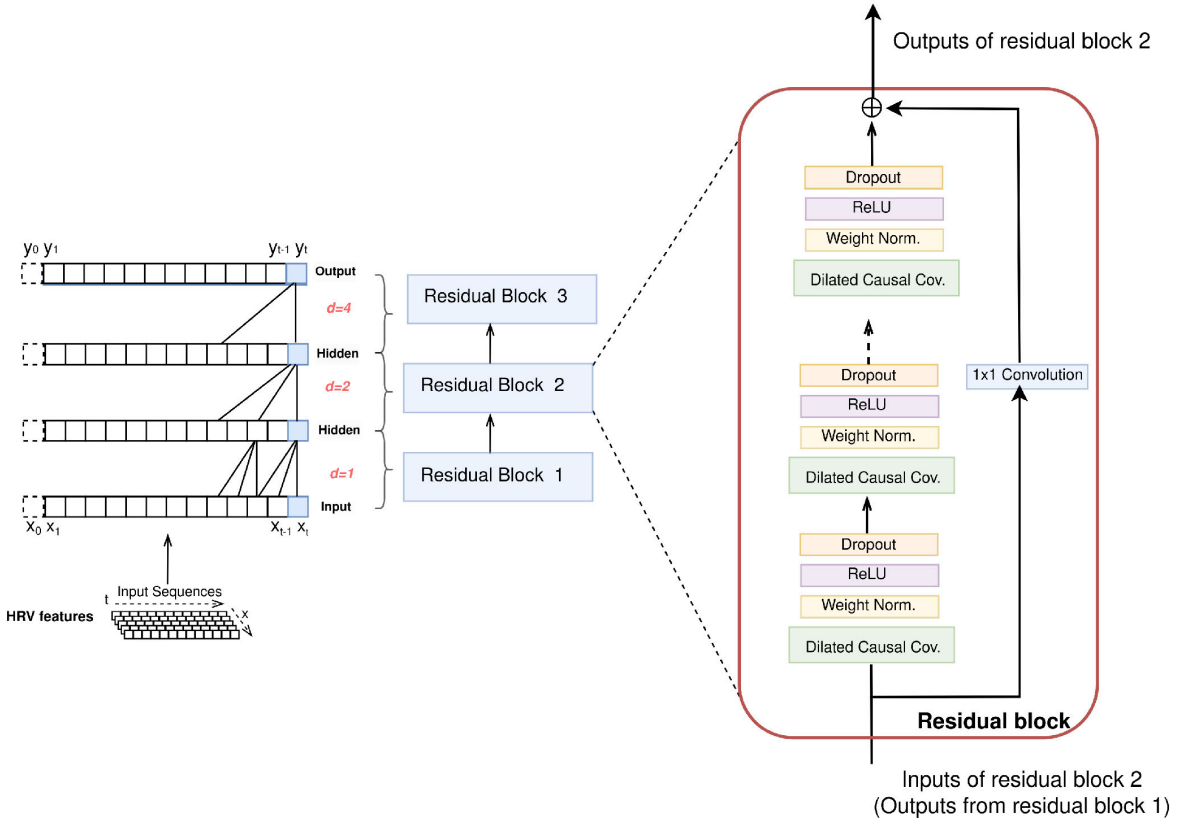

FIGURE 5. This is the block diagram of TCN model. Hourly calculated HRV features are fed to the TCN model and output of the TCN is denoted by y_t. Example for a d-dilated convolution is illustrated on the left hand side with exponential dilation factors $d = 1, 2, 3$ and filter size of 3. This figure is adopted from [36].

HRV measures assess how the heart rate signal changes in inter-beat intervals (IBIs). It commonly relies on three main types of domain namely time, frequency, and non-linear domain. Time-domain measurements demonstrate the total variability of heart rate. Frequency-domain measurements indicate the distribution of spectral power across different frequency bands. Finally, non-linear measurements estimates the unpredictability and complexity of RR series.

Frequency domain related features are calculated based on the two main bands which are high frequencies within a frequency range of 0.15–0.4 Hz band, and low frequencies within a frequency range of 0.04–0.15 Hz band [37].

In order to collect the HRV measurements, we first divide ECG recordings into non-overlapping 5-minute segments. Subsequently, we segment the time series into one-hour-wide bins by calculating the mean of all HRV measurements taken within each 5-minute window. We extract 25 features for the training. Details on the feature description is given in Table 1.

D. SEQUENCE MODELING WITH TEMPORAL CONVOLUTIONAL NETWORKS

This section describes the details of TCN which is a type of CNN architecture used for time series modeling. In [36], it has been adapted to sequence modeling for various types of tasks and dataset while reporting superior performance when comparison to recurrent architectures, such as LSTMs and GRUs. It has been widely utilized to capture dynamic patterns in the trajectory of disease progression [38], [39]. In this study, we employ TCN for its three effective properties:

1) Variable input length: TCN generates an output that matches the length of the input.
2) Causal convolutions: Outputs are only affected by current and previous inputs therefore it emerges as a favorable choice for sequence modeling due to its capability of preventing information leakage from the future to the past.
3) Long-range dependencies: TCN offers flexible receptive size by increasing the dilation factor exponentially with the depth of the network or choosing larger filter sizes. As a result, it maintains large effective history.

TCNs utilize 1D convolutions to capture both local and global temporal patterns in the input data, without relying on recurrent connections as used in recurrent neural networks (RNNs). 1D convolution layer employs causal convolutions, where at a given time t, inputs are limited to present and

past. This is essentially preventing passing information from the future into the past. Additionally, for the cases where sequence length varies, zero-padding is applied to the left side of the input sequence to maintain causality. The problem with the simple causal convolution is that it can only consider a history that is proportional to the depth of the network. Using causal convolutions solely may not be effective on sequences requiring longer history. To accomplish this point, dilated convolutions are proposed [36]. In our settings, we assume where we have N_p patients, each denoted by an index $i \in \{1, 2, .., N_p\}$. For each patient, x^p is a multivariate time series that signifies the observations of HRV variables over time for patient p. A dilated causal convolution with a filter $h : \{0, 1, 2, .., k-1\}$ is applied over the input as follows.

$$F(t) = (x_v^p *_d h)(t) = \sum_{l=0}^{k-1} h(l) \cdot x_v^p(t - l \cdot d). \quad (1)$$

where $F(.)$ denotes the output, x_v^p represents the observations of HRV variable v over time for patient p, $*_d$ denotes the dilated convolution where d is the dilation factor, k is the filter size. The expression $(t - l.d)$ demonstrates the direction towards the past, and it ensures that the network output at time t remains independent of any future information, specifically, HRV observations at times $t + 1, t + 2, \ldots, N_t$. When $d = 1$, it is simply a regular convolution. Increasing d enables a longer effective memory. It was suggested that one can increase d exponentially such that $d = 2^n$ at level n of the network [36]. Details are given in Figure 5.

Furthermore, we structured the TCN with residual blocks, where each temporal block comprises a sequence of operations (including causal convolutions, activation, normalization, and dropout) that are applied to the input data. The output of each block is then combined with the input of the subsequent residual block. Figure 5 illustrates the details of the layers used in the model.

E. EXPERIMENTAL SETUP AND IMPLEMENTATION DETAILS

This section provides more details about the experimental setup and model design after the elimination and the cleaning process. We first start with imputing the time series by implementing a carry-forward imputation. We then employ an imputation scheme if the missing bins are encountered due to the filtering process of non-ECG segments. Finally, if there are still empty bins we apply mean imputation for each patient. It is important to note that no augmentation technique is used. 5-fold stratified cross validation is used to train TCN wherein data is randomly split using 80% of the data for training, 10% of the data for validation, and 10% of the data for testing. For each split, the specific training set is processed to normalize each channel of the time series using z-score normalization. In our experiment, we employ the zero-padding technique to ensure equal layer sizes across the model. This allows variable-length data to be encoded into a single 3-D tensor. (samples × features × times). The time dimension corresponds to the length of the longest time series. Particularly, shorter time series are padded with 0 until reaching the same length as the longest time series in the population.

TABLE 2. Demographic statistic of the final cohort.

Patient Demographics	Sepsis Group	Control Group
Number of patients	65	610
Female	33	288
Male	32	322
Age(μ)	79.62	76.25
Total sepsis hours	17.4h	-
Total ICU length of stay	193.6	84.8h
Ethnicity		
White	43	413
Black or African-American	5	63
Hispanic or Latino	1	26
Unknown	10	62
Other	6	46
Admission Type		
Elective	7	112
Emergency	58	493
Urgent	-	5

TABLE 3. Details of the hyperparameter search.

Hyperparameter	Bounds	Sampling Dist.	Optimum
learning rate	[1e-4,1e-3]	log uniform	5e-4
filter size	[3,8]	uniform	5
residual blocks	[1,8]	uniform	4
batch size	(8,16,32,64)	not applicable	8
hidden layers	[20,200]	uniform	150
weight decay	[0.1, 0.9]	uniform	0.4

For the hyperparameter tuning, we employ grid search on 10% of validation data. Grid search is an exhaustive and expensive method. Therefore, we scale the hyperparameter tuning process by integrating the Ray-Tune framework [40]. The hyperparameters of the TCN model is determined through a random search on 500 runs. The Async Successive Halving Algorithm (ASHA) is chosen for early stopping. Additionally, hyperparameters and checkpoints are determined to maximize the sensitivity. After selecting the best model parameters, they are used to evaluate the test splits.

In the following sections, the model performance will be evaluated based on two main metrics, namely area under the receiver-operating-characteristic curve (AUROC) and area under the precision-recall curve (AUPRC). We emphasize that we will focus on AUPRC results due to the considerable class imbalance.

III. RESULTS AND DISCUSSION

Table 2 outlines the final cohort demographics. After the patient elimination, the final cohort involves 65 sepsis, 650 control patients. It is worth noting that the ICU length of stay of sepsis patients is 2.3 times longer than that of

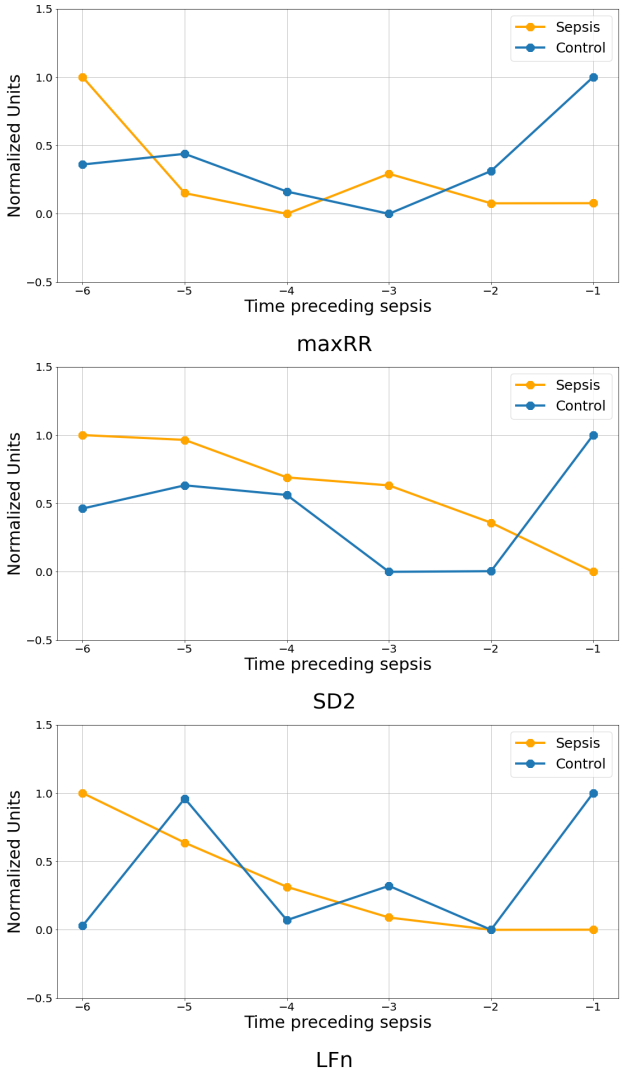

FIGURE 6. The evaluation of median values in sepsis patients compared to control patients during the 6-hour period leading up to sepsis onset.

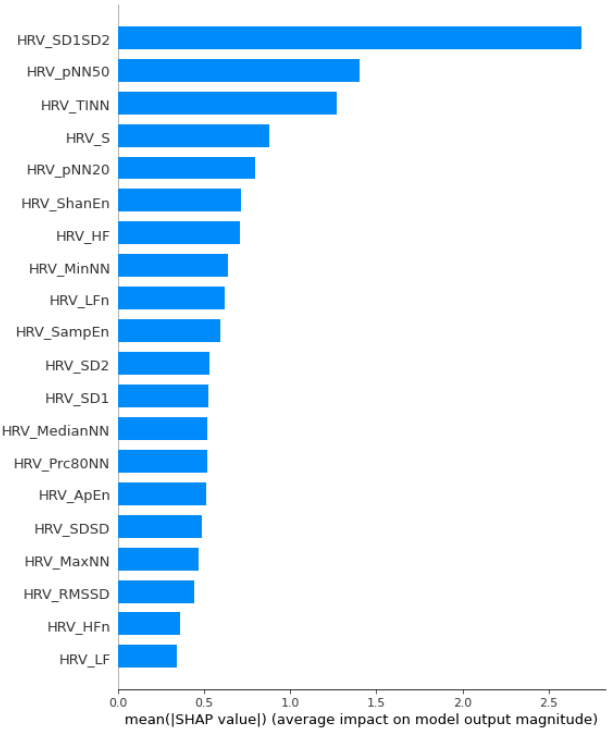

FIGURE 7. Feature importance (top 20) in test dataset based on SHAP analysis.

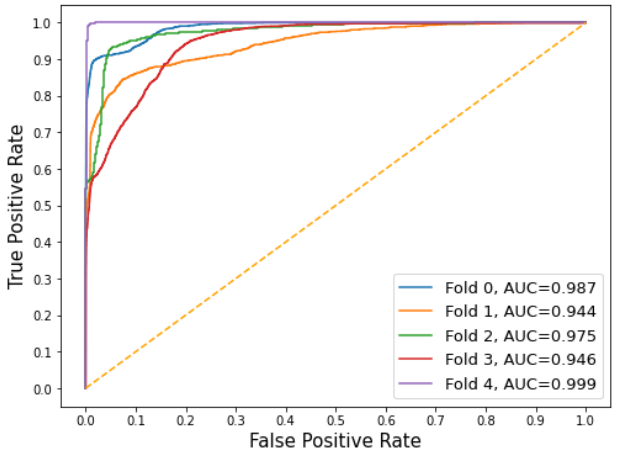

FIGURE 8. ROC curves for the 5 folds obtained from TCN model predicting sepsis onset time.

the control patients. Hyperparameter search is done for five model parameters, namely learning rate, filter size, batch size, momentum and number of hidden layers. Optimal hyperparameters and corresponding search space are reported in Table 3.

In Figure 6, we provide a comparison between the median values of sepsis and control patients for the selected HRV measurements. The top one depicts changes in the median value for the maxRR 6 hours prior to sepsis including both patient groups. It can be observed that as sepsis onset approaches, the median value of maxRR tends to be consistently lower for sepsis patients. Furthermore, it is worth noting that the median value of SD2 is declining among sepsis patients, which corresponds to the reduced HRV associated with sepsis. Similarly, LFn indicates evidence of decreased HRV in sepsis patients when compared to control groups.

It is critical to investigate the correlation and contributions of individual variables to the overall predictions in applications of decision support systems. To quantify and attribute the impact of individual features on model predictions, we calculate Shapley values by using DeepExplainer method. Positive SHAP values are often used to highlight which features are driving the prediction in a positive direction (sepsis). We present the mean absolute Shapley values, which are calculated and averaged on test data. Mean absolute Shapley values are plotted in Figure 7. Notably, it illustrates that SD1/SD2 and pnn50 are the primary contributors to the model's predictions. This also aligns with the results

TABLE 4. The performance of the TCN model for identifying sepsis while testing different time horizons.

Prediction Windows	AUROC %	AUPRC%
Onset time	94.09	82.86
2h prior to onset	83.87	66.54
4h prior to onset	84.83	70.54
6h prior to onset	81.88	71.34

depicted in 6, as we can see a gradual decrease in the SD2 feature. Nonetheless, we strongly recommend conducting additional research into the calculation of Shapley values for time-series data that has been zero-padded and varies in length.

First, we display the ROC curve collected from 5-fold cross validation for sepsis onset prediction in Figure 8. Table 4 presents the prediction performance of the TCN model in our experimental setup. Note that the performance metrics are reported on the test dataset (10%). As it is expected, we received the best performance at the onset time of sepsis with an AUC of 0.941 and 0.828 AUPRC. Specifically, for early prediction tasks of 4h and 6h before sepsis onset, our proposed method yields promising results with AUROC and AUPRC scores of 0.818 and 0.713, respectively. We expect to see gradual performance degradation with increasing distance to the sepsis onset time. However, interestingly 2h prior to onset does not show competitive results. While we compare our findings with the recent research, we notice that keeping the original sepsis to control ratio (10%) makes the classification task considerably challenging. Authors in [41] use the same database and report AUROC and AUPRC scores of 0.92 and 0.90 respectively for identifying sepsis within the first hour of admission. Note that authors [41] are only using 1h recordings in the first hour of intensive care stay and create a balanced population involving 71 sepsis and 71 control patients. The results may be optimistic given the fact that clinical datasets are highly imbalanced in nature. Similarly, they annotate the data based on using Sepsis-3 criteria. However, the details of internally developed signal annotator is not shared. In [24], authors utilize raw PPG signals in the ResNet algorithm resulting in an AUC of 0.84 to classify sepsis at onset time. Our results show superior performance when compared with the results in [24]. In addition, authors simply employ ICD-9 codes for the subject selection. Considering the circumstances where multiple recordings are collected for a particular ICU stay, using ICD-9 codes could cause potential labeling problems without knowing the sepsis onset time. The best practise would be using Sepsis-3 criteria to identify the subjects with sepsis. Finally, we compare our results with the Multi Branch TCN (MB-TCN) model, where the PhysioNet/Computing in Cardiology Challenge 2019 data is used [13]. Authors announce the best AUROC and AUPRC scores of 0.892 and 0.527, respectively, for predicting 6 hour prior to onset. Note that MB-TCN paper is putting effort on balancing the training dataset and teaching model to missingness patterns. We note that the clinical datasets tend to be irregularly-sampled. Further, it is restricted to use for continuous monitoring for early sepsis prediction.

A primary drawback in our study is the limited quantity of sepsis cases. In spite of employing feature selection, cross-validation, and early stopping techniques to prevent overfitting, we notice that the TCN model tends to exhibit signs of overfitting when we exceed 30 epochs. Additionally, we notice early convergence behavior with the TCN model. This encourages us to validate the proposed approach more extensively using a larger dataset or to enhance our signal elimination method in order to recover more data rather than strictly eliminating it. Furthermore, we strongly recommend external validation of the proposed model using data from various medical centers to ensure its applicability across different settings. This concern has been raised in a recent review paper which emphasizes the necessity for multicenter studies to address the challenges related to data size, result validation, resource utilization, and the inclusion of diverse patient population. However, it is important to note that multicenter studies also come with challenges, such as coordinating efforts across different sites, managing data collection and analysis, and addressing variations in local practices such as therapeutic policies. Another constraint of our study is that we exclusively analyze the first 48-hour recordings of patients in the control group. In [42], it was reported that a significant decrease in predictive performance of their initial work was observed when they made a minor modification by implementing case-control matching. The proposed case-control alignment in [23] is an effective technique to explore greater detail.

IV. CONCLUSION

This study established the groundwork for a more comprehensive data integration approach that can help clinicians in their decision-making procedures for sepsis prediction. The motivation behind this work stemmed primarily from our interest in non-invasive and continuous early prediction of sepsis. Consequently, we introduced a systematic workflow aimed at automatically predicting sepsis in long-term ICU ECG recordings sourced from the MIMIC-III waveform database. As the primary result of this study, we observed a gradual decrease in HRV features in the hours preceding the onset of sepsis. Furthermore, especially for the early detection task of 6h before sepsis onset shows promising results with AUROC and AUPRC scores of 0.818 and 0.713, respectively. This confirms that features derived by the ECG models may play an important role in continuous early sepsis prediction.

REFERENCES

[1] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, and C. M. Coopersmith, "The third international consensus definitions for sepsis and septic shock (sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801–810, 2016.

[2] K. E. Rudd, S. C. Johnson, K. M. Agesa, K. A. Shackelford, D. Tsoi, D. R. Kievlan, D. V. Colombara, K. S. Ikuta, N. Kissoon, and S. Finfer, "Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the global burden of disease study," *Lancet*, vol. 395, no. 10219, pp. 200–211, 2020.

[3] S. L. Weiss, J. C. Fitzgerald, F. Balamuth, E. R. Alpern, J. Lavelle, M. Chilutti, R. Grundmeier, V. M. Nadkarni, and N. J. Thomas, "Delayed antimicrobial therapy increases mortality and organ dysfunction duration in pediatric sepsis," *Crit. Care Med.*, vol. 42, no. 11, pp. 2409–2417, Nov. 2014.

[4] World Health Organization. (2018). *Antimicrobial Resistance*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

[5] L. M. Fleuren, T. L. T. Klausch, C. L. Zwager, L. J. Schoonmade, T. Guo, L. F. Roggeveen, E. L. Swart, A. R. J. Girbes, P. Thoral, A. Ercole, M. Hoogendoorn, and P. W. G. Elbers, "Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy," *Intensive Care Med.*, vol. 46, no. 3, pp. 383–400, Mar. 2020.

[6] 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," *Crit. Care Med.*, vol. 46, no. 4, pp. 547–553, 2018.

[7] S. P. Shashikumar, Q. Li, G. D. Clifford, and S. Nemati, "Multiscale network representation of physiological time series for early prediction of sepsis," *Physiolog. Meas.*, vol. 38, no. 12, pp. 2235–2248, Nov. 2017.

[8] S. M. Lauritsen, M. E. Kalør, E. L. Kongsgaard, K. M. Lauritsen, M. J. Jørgensen, J. Lange, and B. Thiesson, "Early detection of sepsis utilizing deep learning on electronic health record event sequences," *Artif. Intell. Med.*, vol. 104, Apr. 2020, Art. no. 101820.

[9] Q. Mao, M. Jay, J. L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Y. Zhou, and R. 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, Jan. 2018, Art. no. e017833.

[10] M. A. Reyna, C. Josef, S. Seyedi, R. Jeter, S. P. Shashikumar, M. B. Westover, A. Sharma, S. Nemati, and G. D. Clifford, "Early prediction of sepsis from clinical data: The PhysioNet/computing in cardiology challenge 2019," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2019, pp. 1–4.

[11] J. Morrill, A. Kormilitzin, A. Nevado-Holgado, S. Swaminathan, S. Howison, and T. Lyons, "The signature-based model for early detection of sepsis from electronic health records in the intensive care unit," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2019, pp. 1–4.

[12] M. Apalak and K. Kiasaleh, "Improving sepsis prediction performance using conditional recurrent adversarial networks," *IEEE Access*, vol. 10, pp. 134466–134476, 2022.

[13] Z. Wang and B. Yao, "Multi-branching temporal convolutional network for sepsis prediction," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 2, pp. 876–887, Feb. 2022.

[14] M. Zabihi, S. Kiranyaz, and M. Gabbouj, "Sepsis prediction in intensive care unit using ensemble of XGboost models," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2019, pp. 1–4.

[15] J. Singh, K. Oshiro, R. Krishnan, M. Sato, T. Ohkuma, and N. Kato, "Utilizing informative missingness for early prediction of sepsis," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2019, pp. 1–4.

[16] B. Rim, N.-J. Sung, S. Min, and M. Hong, "Deep learning in physiological signal data: A survey," *Sensors*, vol. 20, no. 4, p. 969, Feb. 2020.

[17] M. Mollura, G. Mantoan, S. Romano, L.-W. Lehman, R. G. Mark, and R. Barbieri, "The role of waveform monitoring in sepsis identification within the first hour of intensive care unit stay," in *Proc. 11th Conf. Eur. Study Group Cardiovascular Oscillations (ESGCO)*, Jul. 2020, pp. 1–2.

[18] F. M. de Castilho, A. L. P. Ribeiro, V. Nobre, G. Barros, and M. R. de Sousa, "Heart rate variability as predictor of mortality in sepsis: A systematic review," *PLoS ONE*, vol. 13, no. 9, Sep. 2018, Art. no. e0203487.

[19] R. Joshi, D. Kommers, L. Oosterwijk, L. Feijs, C. van Pul, and P. Andriessen, "Predicting neonatal sepsis using features of heart rate variability, respiratory characteristics, and ECG-derived estimates of infant motion," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 681–692, Mar. 2020.

[20] C. León, G. Carrault, P. Pladys, and A. Beuchée, "Early detection of late onset sepsis in premature infants using visibility graph analysis of heart rate variability," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1006–1017, Apr. 2021.

[21] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398.

[22] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, "Convolutional neural networks for patient-specific ECG classification," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2608–2611.

[23] M. Moor, M. Horn, B. Rieck, D. Roqueiro, and K. Borgwardt, "Early recognition of sepsis with Gaussian process temporal convolutional networks and dynamic time warping," in *Proc. Mach. Learn. Healthcare Conf.*, 2019, pp. 2–26.

[24] S. Lombardi, P. Partanen, P. Francia, I. Calamai, R. Deodati, M. Luchini, R. Spina, and L. Bocchi, "Classifying sepsis from photoplethysmography," *Health Inf. Sci. Syst.*, vol. 10, no. 1, p. 30, Oct. 2022.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] J.-M. Kwon, Y. R. Lee, M.-S. Jung, Y.-J. Lee, Y.-Y. Jo, D.-Y. Kang, S. Y. Lee, Y.-H. Cho, J.-H. Shin, J.-H. Ban, and K.-H. Kim, "Deep-learning model for screening sepsis using electrocardiography," *Scandin. J. Trauma, Resuscitation Emergency Med.*, vol. 29, no. 1, pp. 1–12, Dec. 2021.

[27] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W.-H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, May 2016.

[28] R. C. Bone, R. A. Balk, F. B. Cerra, R. P. Dellinger, A. M. Fein, W. A. Knaus, R. M. H. Schein, and W. J. Sibbald, "Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis," *Chest*, vol. 101, no. 6, pp. 1644–1655, Jun. 1992.

[29] *National Early Warning Score (NEWS): Standardising the Assessment of Acute-Illness Severity in the NHS*, London RCoP, London, U.K., 2012.

[30] C. Stenhouse, S. Coates, M. Tivey, P. Allsop, and T. Parker, "Prospective evaluation of a modified early warning score to aid earlier detection of patients developing critical illness on a general surgical ward," *Brit. J. Anaesthesia*, vol. 84, no. 5, p. 663P, May 2000.

[31] C. W. Seymour, V. X. Liu, T. J. Iwashyna, F. M. Brunkhorst, T. D. Rea, A. Scherag, and G. Rubenfeld, "Assessment of clinical criteria for sepsis: For the third international consensus definitions for sepsis and septic shock (sepsis-3)," *J. Amer. Med. Assoc.*, vol. 315, no. 8, pp. 762–774, 2016, doi: 10.1001/jama.2016.0288.

[32] S. D. Subramaniam and B. Dass, "An efficient convolutional neural network for acute pain recognition using HRV features," in *Proc. Int. E-Conf. Intell. Syst. Signal Process.* Cham, Switzerland: Springer, 2022, pp. 119–132.

[33] K. Hovsepian, M. Al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "CStress: Towards a gold standard for continuous stress assessment in the mobile environment," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2015, pp. 493–504.

[34] U. Satija, B. Ramkumar, and M. S. Manikandan, "Automated ECG noise detection and classification system for unsupervised healthcare monitoring," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 722–732, May 2018.

[35] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 3, pp. 832–838, May 2015.

[36] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[37] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers Public Health*, vol. 5, p. 258, Sep. 2017.

[38] D. Jarrett, J. Yoon, and M. van der Schaar, "Dynamic prediction in clinical survival analysis using temporal convolutional networks," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 424–436, Feb. 2020.

[39] F. J. R. Catling and A. H. Wolff, "Temporal convolutional networks allow early prediction of events in critical care," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 3, pp. 355–365, Mar. 2020.
[40] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, "Tune: A research platform for distributed model selection and training," 2018, *arXiv:1807.05118*.
[41] M. Mollura, L.-W.-H. Lehman, R. G. Mark, and R. Barbieri, "A novel artificial intelligence based intensive care unit monitoring system: Using physiological waveforms to identify sepsis," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 379, no. 2212, Dec. 2021, Art. no. 20200252.
[42] J. Futoma, S. Hariharan, K. Heller, M. Sendak, N. Brajer, M. Clement, A. Bedoya, and C. O'Brien, "An improved multi-output Gaussian process rnn with real-time validation for early sepsis detection," in *Proc. Mach. Learn. Healthcare Conf.*, 2017, pp. 243–254.

MERVE APALAK (Member, IEEE) received the B.Sc. degree in electrical engineering from Erciyes University, Kayseri, Turkey, in 2014, and the master's degree in electrical engineering from The University of Texas at Dallas, Richardson, USA, in 2018, where she is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department. She is a Computational Scientist with the Bioinformatics Department, UT Southwestern Medical Center. Her current research interests include clinical data analysis, biological data, transformer models, generative learning, parallel processing, and high-performance computing.

KAMRAN KIASALEH (Senior Member, IEEE) received the B.S. (cum laude), M.S., and Ph.D. degrees in electrical engineering from the Communications Sciences Institute, University of Southern California (USC), Los Angeles, CA, USA, in 1981, 1982, and 1986, respectively. From December 1996 to December 1997, he was on a special assignment with the DSP Research and Development Center, Texas Instruments Inc., Dallas, TX, USA, where he conducted research on various aspects of the third-generation wireless communication systems. From September 2000 to January 2003, he was with Optical Crossing Inc. (OCI), where he launched the development of state-of-the-art free-space and millimeter-wave communications systems. He is currently a Full Professor with the Department of Electrical Engineering, The University of Texas at Dallas (UTD). He has also been the Principle Investigator of a grant from the National Institute of Justice (NIJ) for the development of software-defined radios (SDR) for emergency responders in the USA. His research interests include optical and RF beam propagation through turbulent/unknown media, application of RF technology in cancer detection, tissue optics, synchronization, SDR, and optical communication systems. He was a recipient of the Research Initiation Award (RIA) from the National Science Foundation (NSF). He was also a recipient of the NASA/ASEE Faculty Fellowship Award at the Jet Propulsion Laboratory (JPL), in 1992, where he participated in the Galileo Optical Experiment (GOPEX) demonstration, the first successful demonstration of an optical communications link involving a deep-space vehicle. For his participation in this project, he received the NASA Group Achievement Award. In 1993, he was a recipient of the NASA/ASEE Faculty Fellowship Award at JPL, where he participated in the Compensated Earth-Moon-Earth Laser Link (CEMERLL) demonstration.