Mathematical Biosciences and Engineering

MBE, 19(10): 9825−9841.
DOI: 10.3934/mbe.2022457
Received: 30 April 2022
Revised: 13 June 2022
Accepted: 27 June 2022
Published: 08 July 2022

http://www.aimspress.com/journal/MBE

Research article

Explainable and efficient deep early warning system for cardiac arrest prediction from electronic health records

Qinhua Tang, Xingxing Cen* and Changqing Pan*

Shanghai Chest Hospital, Shanghai Jiaotong University, 241 West Huaihai Road, Shanghai, China

* **Correspondence:** Email: cxx2347@163.com, panchangqing@shchest.org.

Abstract: Cardiac arrest (CA) is a fatal acute event. The development of new CA early warning system based on time series of vital signs from electronic health records (EHR) has great potential to reduce CA damage. In this process, recursive architecture-based deep learning, as a powerful tool for time series data processing, enables automatically extract features from various monitoring clinical parameters and to further improve the performance for acute critical illness prediction. However, the unexplainable nature and excessive time caused by black box structure with poor parallelism are the limitations of its development, especially in the CA clinical application with strict requirement of emergency treatment and low hidden dangers. In this study, we present an explainable and efficient deep early warning system for CA prediction, which features are captured by an efficient temporal convolutional network (TCN) on EHR clinical parameters sequence and explained by deep Taylor decomposition (DTD) theoretical framework. To demonstrate the feasibility of our method and further evaluate its performance, prediction and explanation experiments were performed. Experimental results show that our method achieves superior CA prediction accuracy compared with standard national early warning score (NEWS), in terms of overall AUROC (0.850 Vs. 0.476) and F1-Score (0.750 Vs. 0.450). Furthermore, our method improves the interpretability and efficiency of deep learning-based CA early warning system. It provides the relevance of prediction results for each clinical parameter and about 1.7 times speed enhancement for system calculation compared with the long short-term memory network.

Keywords: early warning system; temporal convolutional network; electronic health records; cardiac arrest; deep taylor decomposition; deep learning

1. Introduction

Cardiac arrest (CA) is a fatal acute event caused by almost any heart disease, such as heart's electrical system malfunctions and cardiomyopathy [1]. Complicated with abrupt loss of heart function and failure of respiratory, this illness causes 58−61% mortality rate in admitted CA patients and even 25−75% long-term side effects proportion in survivors [2]. To reduce the damage of this event, the necessity of CA preventing is emphasized by European Resuscitation Council guidelines, and the early identification of patients at CA risk is considered an effective measure [3]. Traditional studies, such as national early warning score (NEWS) and modified early warning score (MEWS), commonly evaluate the warning score for each vital-sign to quickly determine patient's degree of CA risk [4,5]. However, these methods make each evaluated variables independent and identically distributed, which loses the spatiotemporal information from the vital signs and makes the CA prediction incomplete [6]. Hence, the efficient, precise, and complex data pattern analysis-supported methods are being constantly required for the identification of patients at CA risk.

Currently, deep learning (DL) is the state-of-the-art method used in clinical field to predict adverse events a few hours before their occurrence [6–8]. DL can automatically extract features from various monitoring parameters usually in electronic health records (EHR) and can achieve higher accuracy, by using a neural network structure and a propagation train mechanism [9,10]. For model construction, recursive architectures that are naturally suitable for temporal data processing, such as recurrent neural networks (RNN), long short-term memory (LSTM) and gated recurrent unit (GRU), are widely used to extract event-related features from parameter sequences [11,12]. Applying a RNN-based deep early warning system, Kwon et al.[7] and Hong et al. [13] achieved up to 24.3 and 23.6% higher sensitivity for detection of CA patients compared with a traditional MEWS, respectively. Notably, Kim et al. [8] and Lee et al. [14] proposed an LSTM-based method to predict CA from 1 to 6 h before event, and obtained 20.2 and 15.2% AUROC improvement compared to NEWS, respectively. Moreover, Park et al. [15] applied Bi-LSTM based method, which achieved up to 20.0% higher AUROC for CA prediction compared with a pediatric early warning score. However, the prediction of CA by current DL methods is usually unexplainable, which cannot give an intellectual link between clinical signs parameters and predicted outcomes. Even if the CA alarm is accurately sounded, medical staff is not clear on which indicator to take corresponding action. Revalidation of alarm-related bed sign parameters will take time and adversely affect the effectiveness of CA interventions. On the other hand, the recursion-based algorithm is also an time-consuming method, limited by its inability to perform massively parallel processing, which brings an adverse impact on the dynamic detection of CA and the rapid converge of model training [16].

With advantages of simple calculation procedures and efficient parallel nature, the temporal convolutional network (TCN) maybe a good alternative model to further improve the performance of DL for CA events prediction [17]. The TCN is based on convolutional network, which can be viewed as a combination of convolutional and recurrent architectures, and has been successfully applied in many other clinical acute events prediction, such as the sepsis and kidney injury [18,19]. In practice, the TCN was found to provide accurate results, and was computationally faster than traditional recursion-based methods [17,20]. However, the research of the TCN for CA prediction is in its infancy. Most of the current study focuses on the other acute critical illness and has not yet been developed for CA event.

In addition, the improvements in transparency and interpretability make the DL algorithms more acceptable to clinicians. Despite the highly value of these methods, the standard DL algorithms including TCN are black box, which makes users unclear on how they arrive at a particular classification decision [21]. Particularly in the clinical field, it hinders clinicians from verifying, explaining and understanding the predictions of the DL system, which has huge hidden dangers [22,23]. In the context of explainable artificial intelligence (XAI), numerous methods have been proposed to improve the interpretation of DL, involving a series of theoretical frameworks such as Shapley values, Taylor expansions and the deep Taylor decomposition (DTD) [24]. Compared with the Shapley value, the recent XAI methods based on the DTD framework have better calculation accuracy and derive a series of layer-wise relevance propagation variant methods [25−27]. For clinical application, the DTD is also combined with TCN, which can be used to perform correlation interpretation while performing temporal events prediction to enhance the clinical translation of DL [28]. Nevertheless, the research of the XAI for CA event prediction is still a new direction. As far as our knowledge, the explainable DL research based on DTD theory framework for CA event prediction has not been reported yet.

In this work, we present an explainable and efficient deep early warning system for CA prediction from EHR. Feeding time series of vital signs from EHR to DL model of TCN for CA event prediction task, this method is expected to improve model calculation efficiency while maintaining well precision. Based on predicted results, the interpretability in our architecture is introduced through DTD theoretical framework. Capturing input features by applying Taylor expansions on each layer neurons, this approach is hoped to establish a more transparent prediction process to improve the clinician acceptance of DL. The major contributions of this work are two-fold: 1) The temporal convolutional network was first used as the CA prediction model through the analysis of time series of vital signs from EHR; 2) The deep Taylor decomposition of theoretical framework is first introduced in temporal convolutional network for interpretability of CA prediction.

2. Methods

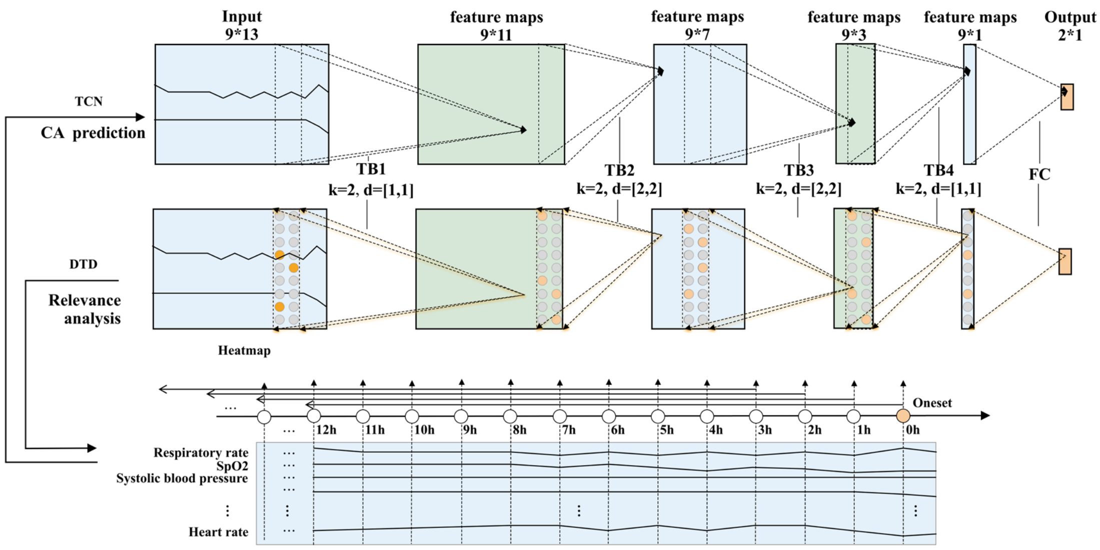

Figure 1. Explainable and efficient deep learning model combined with temporal convolutional network and deep Taylor decomposition for cardiac arrest prediction on electronic health record.

Our deep early warning system consists of two components and is performed as follows. Each selected EHR clinical parameters sequence intercepted with 12-hour window is fed to TCN of deep early warning system, and the probability of the CA event occurrence is predicted through the trained model. Next, these predicted results are back-forward fed to TCN, which combined with DTD explainable theoretical framework to calculate the relevance of clinical parameter to the predicted results. The overall processing flow of our system is shown in Figure 1.

2.1. Cardiac arrest predication model

A TCN-based deep early warning system is introduced in our CA perdiction study, and the computational details of components are described below.

2.1.1. Temporal convolutional network

TCN is the convolutional network-based model which we consider for CA predication. It provides powerful time series processing and maintains a good parallel structure by using causal convolution, dilated convolution, and identity mapping. The concrete architecture is shown in Figure 2. For a given time series input of T length $X = \{x_1, x_2, \ldots, x_T\}$, causal convolution layer is first used to trace their historical characteristics at the each position x_t:

$$(F * X)_{(x_t)} = \sum_{k=1}^{K} f_k x_{t-K+k}, \tag{1}$$

where $F = \{f_1, f_2, \ldots, f_k, \ldots, f_K\}$ is the convolution kernel; K is the size of convolution kernel. Combined with the dilatation scheme, the result of causal convolution contains a larger feature capture range without causing the information loss like pooling layer. With dilatation rate of d, the dilated causal convolution result can be calculated as follows:

$$(F *_d X)_{(x_t)} = \sum_{k=1}^{K} f_k x_{t-(K-k)d}, \tag{2}$$

As a convolutional network-based DL model, the TCN can extract richer features at different levels through the layers' increase. The identity mapping is thus added to the result, which is processed by two rounds of dilated causal convolution, ReLU and Dropout layers, to reduce the influence of model overfitting and network degradation when the number of model layers' increases. Considering the layers transformations F, the identity mapped temporal block of the TCN can be expressed as follows:

$$H(x) = F(x) + x. \tag{3}$$

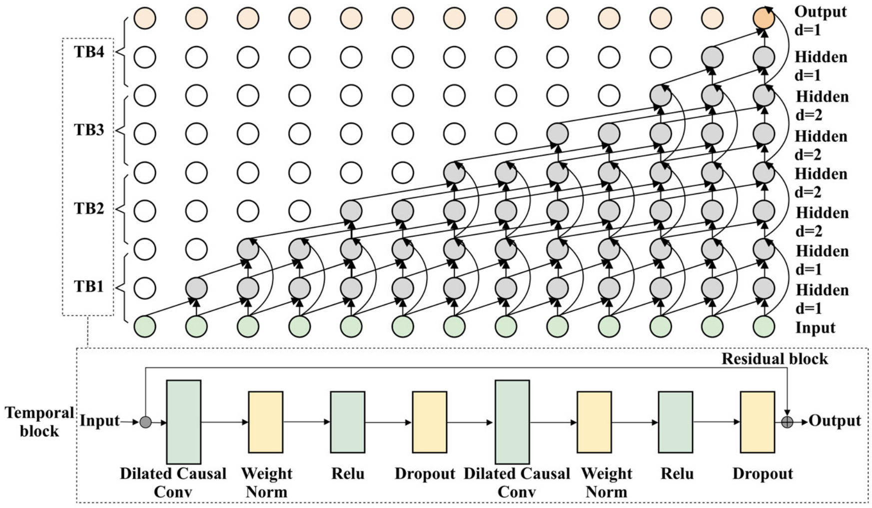

Figure 2. Architecture of temporal convolutional network for cardiac arrest prediction with causal convolution, dilated convolution and residual block.

2.1.2. Temporal convolutional network for cardiac arrest prediction

TCN is followed by fully connected layer with sigmoid function. Sigmoid map the output of multiple neurons (temporal block) to the interval [0, 1], which can be regarded as the probability of being classified into a certain category. For the T length input $X = \{x_1, x_2, \ldots, x_T\}$ and the binary output, the result of the fully connected layer with Sigmoid score is as follows:

$$y_j = \mathrm{sigmoid}(z_j) = \mathrm{sigmoid}(W_j \cdot X + b_j), \tag{4}$$

where W, b are the weight and the bias term of the fully connected layer, respectively; the sigmoid score can be defined as:

$$\mathrm{sigmoid}(z_j) = \frac{1}{1+e^{-z_j}}. \tag{5}$$

2.2. Cardiac arrest explanation model

2.2.1. Deep Taylor decomposition

DTD is an explainable theoretical framework, which is used to recursively decomposes the network output into the contributions of the previous layer network until the input unit [21]. Assuming the layer adjacent neurons u_i, u_j (the u_j is close to the network output), the relevance of the neuron u_j associated with the neuron output x_j can be Taylor decomposed as follows:

$$R_j = \left(\frac{\partial R_j}{\partial\{x_i\}}|_{\{\tilde{x}_i\}^{(j)}}\right)^T \cdot \left(\{x_i\} - \{\tilde{x}_i\}^{(j)}\right) + \varepsilon_j = \sum_{i=1}^{I} \underbrace{\frac{\partial R_j}{\partial x_i}|_{\{\tilde{x}_i\}^{(j)}} . \left(x_i - \tilde{x}_i^{(j)}\right)}_{R_{ij}} + \varepsilon_j, \tag{6}$$

where ε_j is the Taylor residual; R_{ij} identified the redistributed relevance from neuron u_i close to the network input to u_j; I is the number of neurons in layer i; $|_{\{\tilde{x}_i\}^{(j)}}$ indicates that the derivative has been evaluated at the root point $\{\tilde{x}_i\}^{(j)}$, which satisfy the following condition:

$$H_{i,j}(\{\tilde{x}_i\}^{(j)}) = 0, \tag{7}$$

where $H_{i,j}$ is the network calculation function between layers i and j. On the other hand, through the collection of the relevance contributions from all neurons in the upper layer j, the relevance of the neuron u_i can be defined as follows:

$$R_i = \sum_{j=1}^{J} R_{ij}, \tag{8}$$

where J is the neurons number in layer j. Combining the Eqs (6) and (8), the relevance of the neuron u_i associated with the upper layer can then be presented as follows:

$$R_i = \sum_{j=1}^{J} \frac{\partial R_j}{\partial x_i}|_{\{\tilde{x}\}^{(j)}} \cdot \left(x_i - \tilde{x}_i^{(j)}\right). \tag{9}$$

Through the Eq (9), the relevance result of each neuron about forward output can be back propagated to the input of the network.

2.2.2. Deep Taylor decomposition for temporal convolutional network explanation

Combined with the DTD theoretical framework, the training-free relevance model for TCN explanation is introduced [21]. The training of each neuron is not needed by our model, which saves the cost of CA relevance evaluation. Considering the basic unit of feature extraction as shown in Figure 3.

$$\begin{cases} x_{t,j} = max\left(0, \ \sum_{k=1}^{K}\sum_{i=1}^{I} f_{k,i,j}\left(x_{t-(K-k)d,i} + res_{t,j}(X)\right) + b_i\right) \\ x_{t,n} = max\left(0, \ \sum_{k=1}^{K}\sum_{j=1}^{J} f_{k,j,n}\left(x_{t-(K-k)d,j}\right) + b_j\right) \\ x_{t,p} = \text{sigmoid}\left(x_{t,o}\right) \end{cases}, \tag{10}$$

where subscripts i, j, n, o present adjacent network layers, which contain I, J, N, O neurons number, respectively; $res_{t,n}(X)$ is the residual term which related to the output of the previous layer of the entire network X. In that case, we assume that the upper layer has been explained by z+-rule [21]. For the output unit of $x_{t,p} = \text{sigmoid}\left(x_{t,o}\right)$, the total number of correlations of backpropagation is equivalent to the predicted output: $R_{t,p} = x_{t,p}$, and the relevance $R_{t,o}$ of the neuron u_o can be obtained associated with the output layer neuron u_p as follows (according to Eq (9)):

$$R_{t,o} = \frac{\partial x_{t,p}}{\partial x_{t,o}}|_{\{\tilde{x}_{t,o}\}}\left(x_{t,o} - \tilde{x}_{t,o}\right) = \frac{1}{1+e^{-x_{t,o}}}\left(1 - \frac{1}{1+e^{-x_{t,o}}}\right)|_{\{\tilde{x}_{t,o}\}}\left(x_{t,o} - \tilde{x}_{t,o}\right), \tag{11}$$

where $\tilde{x}_{t,o}$ is the root point of the neuron u_o, which is set to 0 value in our work. For the basic TCN unit of $x_{t,n} = max\left(0, \sum_{k=1}^{K}\sum_{j=1}^{J} f_{k,j,n}\left(x_{t-(K-k)d,j}\right) + b_j\right)$, the relevance R_j of the neuron u_j can be obtained associated with the upper layer neuron u_n as follows:

$$R_{t,j} = \sum_{n=1}^{N} \frac{\sum_{k=1}^{K} x_{t-(K-k)d,j} f_{k,j,n}^{+}}{\sum_{k=1}^{K}\sum_{j'=1}^{J} x_{t-(K-k)d,j'} f_{k,j',n}^{+}} R_{t,n}, \tag{12}$$

where $f_{k,j,n}$ is the weight of dilated causal convolution; $f_{k,j,n}^{+}$ represents the positive part of the $f_{k,j,n}$. For the TCN unit of $x_{t,j} = max\left(0, \sum_{k=1}^{K}\sum_{i=1}^{I} f_{k,i,j}\left(x_{t-(K-k)d,i} + res_{t,j}(X)\right) + b_i\right)$, the relevance R_i of the neuron u_i can be obtained associated with the upper layer neuron u_j as follows:

$$R_{t,i} = \sum_{j=1}^{J} \frac{\sum_{k=1}^{K} (x_{t-(K-k)d,j} + res_{t,i}(R)) f_{k,i,j}^{+}}{\sum_{k=1}^{K}\sum_{i'=1}^{I} x_{t-(K-k)d,i'} f_{k,i',j}^{+}} R_{t,j}, \tag{13}$$

where $res_{t,i}(R)$ is the residual term which related to the relevance R of the previous layer of the entire network. According to the Eqs (11)−(13), the relevance terms of the basic unit are modeled, and the training-free relevance is given.

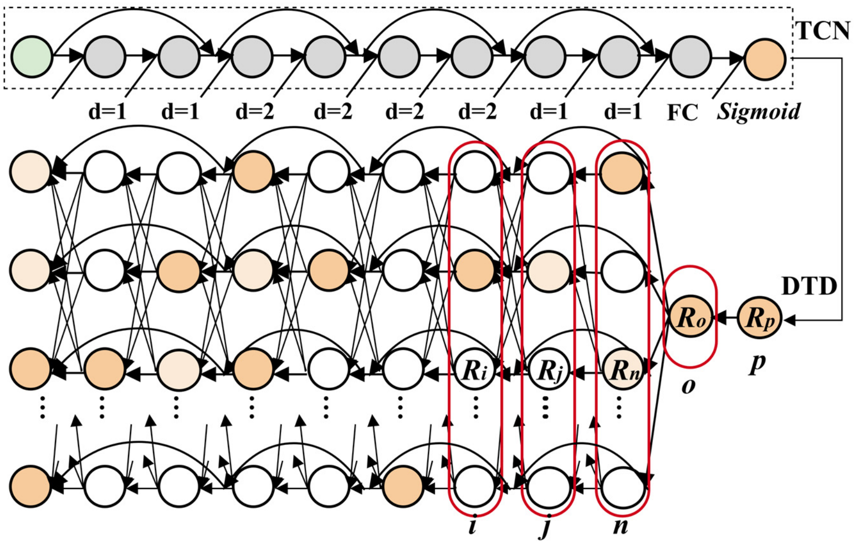

Figure 3. Architecture of temporal convolutional network used deep Taylor decomposition theoretical framework for cardiac arrest explainable prediction.

3. Materials and experiments

3.1. Materials

Our research used electronic health records (EHR) collected from the MIMIC-III, a freely available critical care database [26]. We checked data integrity and randomly selected 486 samples, including 107 CA positive and 379 negative patients. Furthermore, the synthetic minority over sampling technique (SMOTE) was used in our datasets to alleviate the impact of data imbalance on model training.

In detail, the vital-sign measurements in the above records were extracted according to NEWS selection, and basic information of age and sex were also included in our datasets. Each sample contains 7 continuous clinical variables, and the overall of the clinical parameters in our dataset are shown in Table 1.

Table 1. Clinical parameters summary.

Parameters	Unit	CA positive (N = 107)	CA negative (N = 379)
Respiratory rate	bpm	20.9	20.1
SpO2	%	94.7	96.8
Any supplemental oxygen?	%	58.5	62.8
Systolic blood pressure	mmHg	109.3	119.4
Body temperature	°C	36.4	36.8
Heart rate	bpm	91.2	88.1
Conscious Level	-	0.99	0.99
Age	-	68.2	63.0
Male sex	%	61.6	54.4

Table 1 shows the mean value of each selected clinical parameter in different CA categories.

Data pre-processing focuses on the generation of EHR event sequence. Taking the time points of 0, 6, 8, 9, 10, and 11 h before CA event as the end point, the records of each clinical variable were intercepted backward with a 12-hour time window, respectively. Among them, the clinical parameters in time window were each divided into one-hour periods, and the data that occurs in which time period were represented by their average values. On the other hand, through the observation of the experimental datasets, the data missing is seriously. To ensure the data integrity and the subsequent calculations can be properly progressed, generative adversarial nets of AL method was used to imputed missing values. For CA prediction by deep early warning system, we set the ratio of training set, test set and validation set to 3:1:1.

3.2. Experimental setup

In this work, we designed two experiments: 1) prediction experiment compared the CA recognition accuracy from EHR sequence before event occurrence to evaluate the performance of TCN-based early warning system for CA prediction. The F1-Score of the results were calculated, which consisted on classifying EHR sequence belonging to CA positive or negative. 2) explanation experiment presented the back-propagated CA relevance corresponding to clinical parameters to evaluate the interpretability of DTD theoretical framework-based TCN. Mean relevance of each clinical parameter were calculated, which evaluates explainable model from global perspective. In above experiments, the prediction model of TCN and the explainable theoretical framework of DTD were the objects to be verified.

3.2.1. Prediction experiment

In this experiment, the TCN-based CA prediction model was the object to be verified, and the commonly applied early warning system of NEWS, machine learning method of random forest (RF) and the deep learning prediction models of LSTM and GRU were used as the contrast. During the evaluation process, we set up a series of experimental tasks. First, the indicators of sensitivity, specificity, positive value (PPV), negative predictive value (NPV), area under the receiver operating characteristic curve (AUROC), and F1-Score from different models for CA prediction were calculated to evaluate the model feasibility. Then, prediction time and training time of CA prediction from different deep learning models with same layer distribution were calculated to evaluate the model efficiency. Prediction time counts the time it takes for the model to complete CA predictions on a validation dataset. Training time calculates the time it takes for the model to run in a training epoch.

The concrete CA prediction standards of NEWS ware shown in Table 2 [4]. It is considered as a high-risk object of CA when NEWS score is higher than 7 points. For LSTM and GRU, a deep prediction model was constructed by multi-layer LSTM and GRU respectively, which combined with processing of ReLU, dropout, fully connected layer and Sigmoid to obtain the CA prediction result. The models hyperparameters followed the settings in O. Almqvist [20].

For TCN, the overall model architecture and the calculation process are described in Section 2.1. In detail, the TCN was constructed by four temporal blocks, each of which contains two rounds processing of dilated causal convolution, ReLU, dropout successively. The total eight dilated causal convolution layers were composed of [9, 9, 6, 6, 3, 3, 2, 2] size convolution kernels, with dilatation rate of [1, 1, 2, 2, 2, 2, 1, 1], respectively. The architecture configuration of deep learning models were described in Table 3. Training optimization was performed using the Adam algorithm with a batch size of 128 and a learning rate of 1e-3. Furthermore, we set the dropout rate to 0.4 to reduce overfitting of deep model.

The all construction and calculation of CA prediction models were completed on Pytorch using NVIDIA GEFORCE RTX 2080 GPU, and the calculation was terminated without significant changes in training accuracy.

Table 2. National Early Warning Score (NEWS).

NEWS	3	2	1	0	1	2	3
Respiratory rate		≤ 8		9-14	15-20	21-29	≥ 30
SpO2	≤ 91	92-93	94-95	≥ 96			
Any supplemental oxygen?		Yes		No			
Body temperature	≤ 35		35.1-36	36.1-38	38.1-39	≥ 39.1	
Systolic blood pressure	≤ 90	91-100	101-110	111-219			
Heart rate	≤ 40		41-50	51-90	91-110	111-130	≥ 131
Level of consciousness using the AVPU system				A			V, P or U

Note: Level of consciousness: A = Alert; V = Responds to voice; P = Responds to pain; U = Unresponsive.

Table 3. Architecture configuration of deep learning models.

Parameter	LSTM	GRU	Our
Layers	8	8	8
Units per layer	[9, 9, 6, 6, 3, 3, 2, 2]	[9, 9, 6, 6, 3, 3, 2, 2]	[9, 9, 6, 6, 3, 3, 2, 2]
Batch size	128	128	128
Activation Function	ReLU, Sigmoid	ReLU, Sigmoid	ReLU, Sigmoid
Learning rate	1e-3	1e-3	1e-3
Dropout rate	0.4	0.4	0.4
Trainable parameters	2543	1914	750

3.2.2. Deep Taylor decomposition for temporal convolutional network explanation

For TCN explanation, the overall model architecture and the specific calculation process are presented in Section 2.2. Fast training-free DTD method was used to explain the CA prediction made by TCN on the dataset of EHR event sequence. The structure of TCN was maintained, in which the back-propagation calculation based on DTD was performed to obtain the relevance of clinical parameters. For all unit of TCN and FC layers, the z+-rule is applied, which corresponds to the αβ-rule with $\alpha = 1$ and $\beta = 0$ in positive input spaces [27].

4. Results

4.1. Prediction results

In the prediction experiment, the performance of deep early warning system for CA prediction using TCN are shown in Table 4. During this process, the items of sensitvity, specificity, PPV, NPV AUROC and F1-Score were used as the evaluation indicators to validate the model performance. Overall, the prediction results based on our method (AUROC: 0.850, 0.833, 0.801, 0.770; F1-Score: 0.75, 0.694, 0.680, 0.622) are better than those based on NEWS (AUROC: 0.622, 0.579, 0.580, 0.571; F1-Score: 0.476, 0.418, 0.416, 0.399), RF (AUROC: 0.751 0.751, 0.751, 0.751; F1-Score: 0.615, 0.611, 0.564, 0.511) and other DL models of GRU (AUROC: 0.758, 0.751, 0.669, 0.669; F1-Score: 0.631, 0.607, 0.553, 0.507) and LSTM (AUROC: 0.827, 0.767, 0.758, 0.645; F1-Score: 0.698, 0.656, 0.615, 0.537). Figure 4 shows the receiver operating characteristic curves of our method, LSTM, GRU and MEWS, predicting CA event within 1 h.

Table 4. Accuracy comparison for cardiac arrest prediction methods.

Time	Methods	Sensitvity	Specificity	PPV	NPV	AUROC	F1-Score
1 h	NEWS	0.628	0.680	0.383	0.852	0.622	0.476
	RF	0.666	0.869	0.571	0.909	0.751	0.615
	GRU	0.529	0.901	0.782	0.741	0.758	0.631
	LSTM	0.550	0.977	0.956	0.709	0.827	0.698
	Our	0.750	0.902	0.750	0.902	0.850	0.750
2 h	NEWS	0.552	0.656	0.337	0.822	0.579	0.418

Continued on next page

Time	Methods	Sensitvity	Specificity	PPV	NPV	AUROC	F1-Score
2 h	RF	0.733	0.861	0.523	0.939	0.751	0.611
	GRU	0.531	0.870	0.708	0.758	0.751	0.607
	LSTM	0.525	0.933	0.875	0.688	0.767	0.656
	Our	0.708	0.869	0.680	0.883	0.833	0.694
5 h	NEWS	0.533	0.674	0.341	0.820	0.580	0.416
	RF	0.611	0.855	0.523	0.893	0.751	0.564
	GRU	0.439	0.866	0.750	0.629	0.685	0.553
	LSTM	0.500	0.888	0.800	0.666	0.758	0.615
	Our	0.680	0.869	0.680	0.869	0.801	0.680
11 h	NEWS	0.504	0.677	0.331	0.812	0.571	0.399
	RF	0.578	0.808	0.458	0.873	0.751	0.511
	GRU	0.414	0.800	0.653	0.600	0.648	0.507
	LSTM	0.450	0.800	0.666	0.620	0.645	0.537
	Our	0.560	0.901	0.700	0.833	0.770	0.622

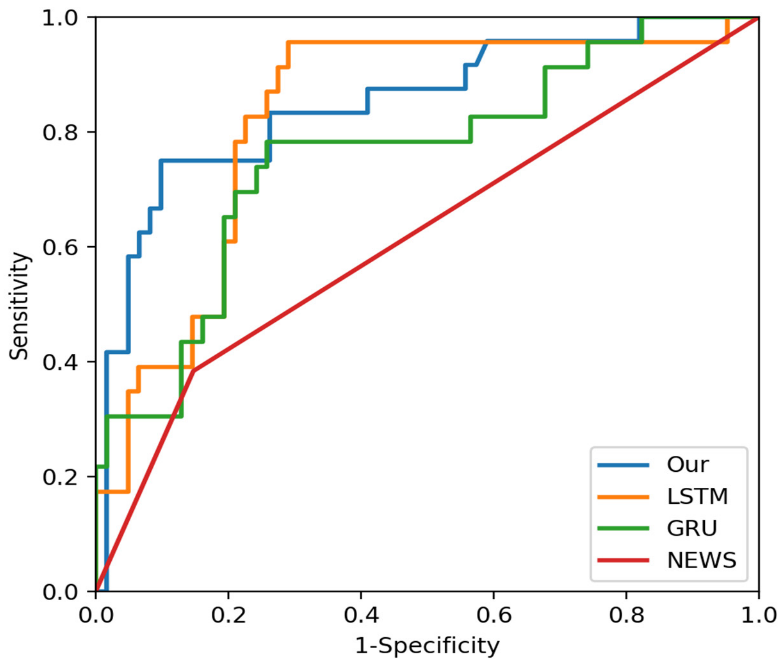

Figure 4. The receiver operating characteristic of methods for detecting cardiac arrest.

For CA detection, the best performance is achieved by our method with sensitivity of 0.750, 0.708, 0.680, 0.560 at time points of 1, 2, 5 and 11 h before the event occurrence, respectively. The best sensitivity improvements of our method compared with NEWS and LSTM have achieved 28.2% (time point of 2 hour before CA event) and 36.3% (time point of 1 hour before CA event), respectively. As expected, the TCN-based deep early warning system can capture more suitable information in EHR clinical parameter sequence to improve the CA prediction.

More detailed accuracy performance of CA early warning system based on different prediction methods are investigated by reporting the changes of F1-Score over time in 12-hour window. Corresponding to the compared NEWS, LSTM and GRU, the performance at each checked time point of CA early warning system is shown in Figure 5. It is observed that the result curves based on TCN are higher than the results based on other models, and the height gap of result curve also maintains stable with the increase of the time interval between the occurrence of CA events. The above results indicate that the deep early warning system based on convolutional units may capture more useful information from the sequence of EHR clinical parameters to improve the performance of CA prediction.

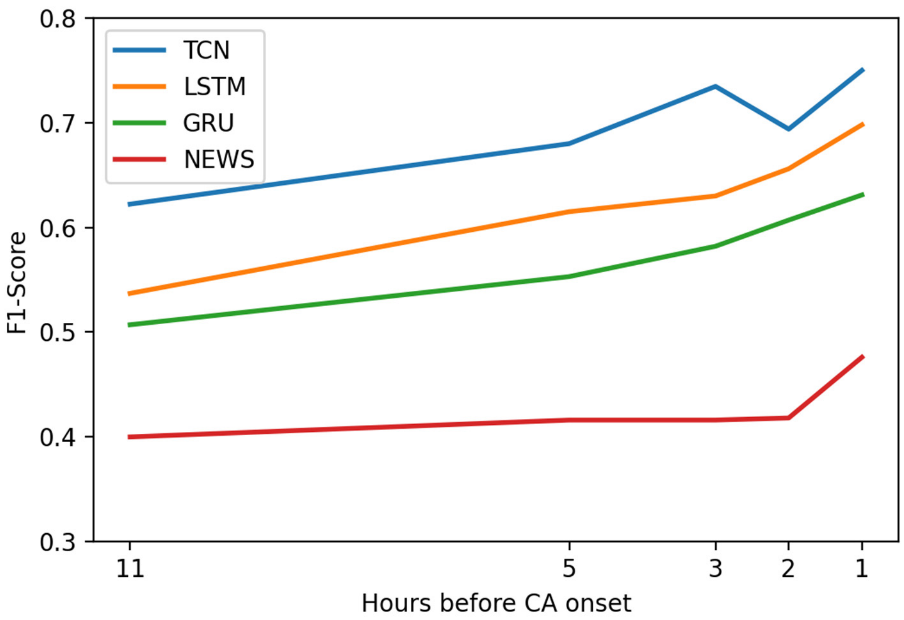

Figure 5. Results of the cardiac arrest prediction experiment within 12-hour window before event occurrence: the prediction results evaluated by F1-Score.

For the model training, it can be seen from Figure 6(a) that the loss change gradually stabilized when the number of training epochs > 3000. In the training epoch interval, all methods tends to converge and the F1-Score based on the TCN is higher than the result of the other deep learning methods, as shown in Figure 6(b). However, the F1-Score of all methods tend to decrease with increasing epoch at the end, which may be due to overfitting of the deep model with many parameters and small training dataset.

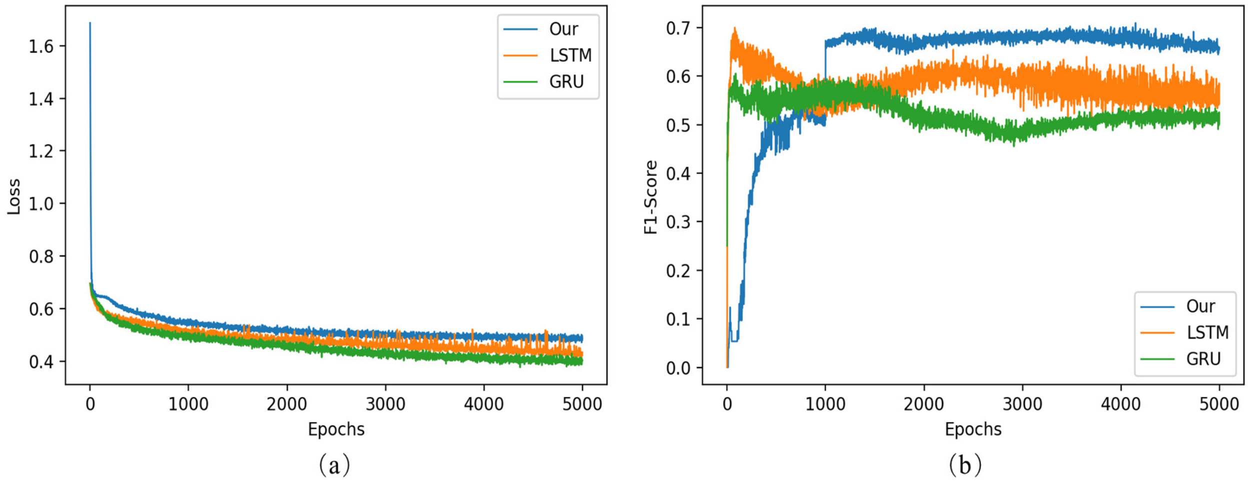

Figure 6. Performance of the deep learning methods for cardiac arrest prediction during model training: (a) the Loss curves with train dataset. (b) the F1-Score curves with validation dataset.

Furthermore, the computational times of the three deep learning models for CA prediction was listed in Table 5. It can be seen that the TCN reduced the CA prediction time consumption by about 76 and 46% compared with the LSTM and GRU respectively. This means that the TCN has huge potential for efficiency improvement in CA prediction.

Table 5. Time computation of the CA prediction models.

Models	Prediction Time (ms)	Training Time (s)
LSTM	0.198	0.270
GRU	0.163	0.265
Our	0.112	0.229

4.2. Explanation results

In the explanation experiment, the performance of TCN for CA explainable prediction using DTD theoretical framework are reported in Figure 7. The mean relevance of each clinical parameter and relevance distribution with different parameter measurements are used as two explainable perspectives of CA prediction. The first explainable perspective points out which clinical parameters in the EHR sequence were relevant to CA prediction. The second explainable perspective points out the details of relevance changes corresponding to the magnitude of clinical parameter, which improved the transparency of deep early warning system for CA prediction and increased the clinicians' understanding of internal CA prediction mechanism of the DL model.

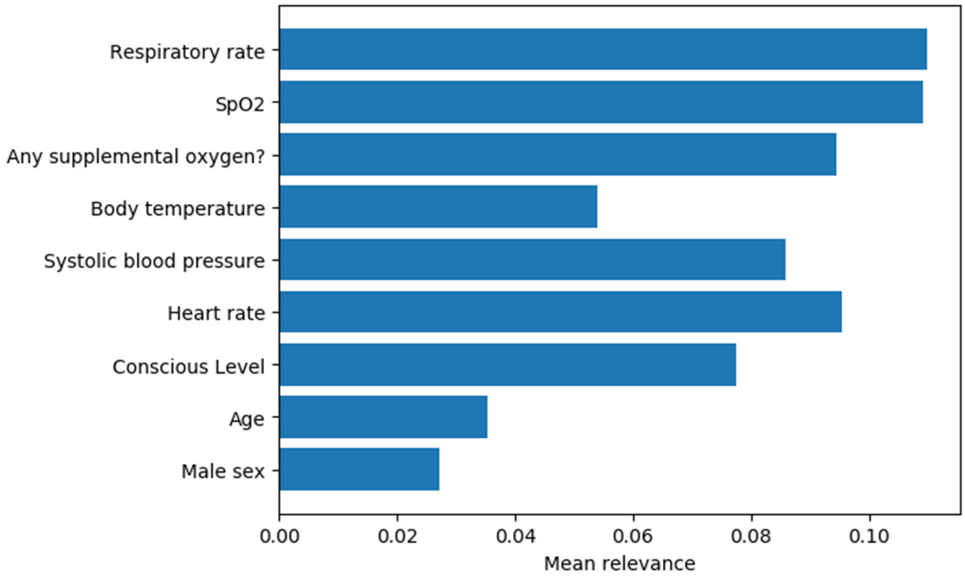

Figure 7. Results of the CA explanation experiment: the mean relevance of each clinical parameter.

In Figure 7, with mean relevance of each clinical parameter, it can be seen that the importance of clinical parameter to CA prediction can be displayed through TCN based on DTD framework. It was observed that the three most important clinical parameters are "Respiratory rate", "SpO2" and "Heart rate" with the corresponding relevance values of 0.11, 0.11 and 0.95, respectively.

5. Discussion and conclusions

We present a TCN-based deep early warning system for CA prediction and apply the DTD theoretical framework to improve the interpretability of the prediction results. Recapitulating previous studies about deep early warning system for CA prediction, these works do achieve the desired effect and enhance the model transparency. Nevertheless, such approaches usually use recursive methods, which cannot perform highly parallel degree accelerated combined with corresponding explainable theoretical framework to better improve the efficiency and interpretability of CA event prediction, whereas the ascendancy of our TCN based approach combined with DTD framework can make up the above deficiencies.

Hitherto, many deep early warning studies of CA prediction have used recursive architecture-based DL method, which may sacrifice or obscure the efficiency nature of CA prediction. Based on structure of convolutional network, TCN captures medical information from EHR sequences with higher parallelism, which is used as a supplement of deep early warning system to increase the performance of DL in CA prediction tasks. When making CA prediction from EHR sequences, we evaluated the capabilities of TCN to capture clinical features and the effectiveness of passing the convolutional network to deep early warning system. We found that proper use of TCN can further ameliorate the performance of CA prediction (see Tables 4, 5 and Figures 4–6). Specifically, it seems can achieve the better performance gains that building the convolution unit to CA prediction. In addition, we also found that explainable theoretical framework based on DTD can further improve the transparency of deep early warning system in the CA explanation task. Indeed, the TCN method combining DTD explainable theoretical framework has advantages over traditional methods in CA prediction. In summary, we found that applying TCN and DTD to CA prediction is a convincing approach, with better performance compared with traditional NEWS and other common DL models of LSTM and GRU.

However, our system still needs to be improved. For CA prediction from EHR clinical parameters sequences, the TCN-based method still has not obtained high-accuracy of prediction result, it may be that the ability of DL method to capture clinical parameters cannot be well released, limited by the insufficient amount of model training data. Meanwhile, the imputed missing values have a certain error with the actual data, which also creates uncertain factors for the model calculation. On the other hand, the model interpretability of TCN based on DTD theoretical framework needs to be further verified, especially the use of quantitative evaluation methods.

Future work would attempt to collect more CA patient samples for the training of the TCN to improve the accuracy and generalization ability of the deep early warning system. Furthermore, a lot of work is required to transform the TCN into CA prediction and then prevent the occurrence of CA event, and the real-time performance of CA prediction would attempt to achieved correspondingly through TCN's parallel computing acceleration optimization. Finally, more exploratory work, such as finding other CA-related clinical parameters, fine-tuning model parameters, medical demonstration of relevance calculation, etc., can allow us to better release the potential of deep CA prediction systems.

Acknowledgments

This work is funded by the Shanghai Municipal Commission of Economy and Information (Grant No. 202002009).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. S. Girotra, B. K. Nallamothu, J. A. Spertus, Y. Li, H. M. Krumholz, P. S. Chan, Trends in survival after in-hospital cardiac arrest, *N. Engl. J. Med.*, **367** (2012), 1912–1920. https://doi.org/10.1016/j.jemermed.2013.02.007
2. L. Mandigers, F. Termorshuizen, N. F. Keizer, D. Gommers, D. Reis Miranda, W. J. Rietdijk, et al., A nationwide overview of 1-year mortality in cardiac arrest patients admitted to intensive care units in the Netherlands between 2010 and 2016, *Resuscitation*, **147** (2020), 88–94. https://doi.org/10.1016/j.resuscitation.2019.12.029
3. J. Soar, J. P. Nolan, B. W. Böttiger, G. D. Perkins, C. Lott, N. I. Nikolaou, et al., European resuscitation council guidelines for resuscitation 2015: section 3. Adult advanced life support, *Resuscitation*, **95** (2015), 100–147. https://doi.org/10.1016/j.resuscitation.2015.07.016
4. G. B. Smith, D. R. Prytherch, P. Meredith, P. E. Schmidt, P. I. Featherstone, The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death, *Resuscitation*, **84 (**2013**)**, 465–470. https://doi.org/10.1016/j.resuscitation.2012.12.016

5. I. Nishijima, S. Oyadomari, S. Maedomari, R. Toma, C. Igei , S. Kobata, et al., Use of a modified early warning score system to reduce the rate of in-hospital cardiac arrest, *J. Intensive Care*, **4** (2016), 1–6. https://doi.org/10.1186/s40560-016-0134-7
6. F. E. Shamout, T. Zhu, P. Sharma, P. J. Watkinson, D. A. Clifton, Deep interpretable early warning system for the detection of clinical deterioration, *IEEE J. Biomed. Heal. Inf.*, **24** (2019), 437–446. https://doi.org/10.1109/jbhi.2019.2937803
7. J. Kwon, Y. Lee, Y. Lee, S. Lee, J. Park, An algorithm based on deep learning for predicting in-hospital cardiac arrest, *J. Am. Heart Assoc.*, **7** (2018), 1−11. https://doi.org/10.1161/jaha.118.008678
8. J. Kim, M. Chae, H. J. Chang, Y. A. Kim, E. Park, Predicting cardiac arrest and respiratory failure using feasible artificial intelligence with simple trajectories of patient data, *J. Clin. Med.*, **8** (2019), 1336−1350. https://doi.org/10.3390/jcm8091336
9. T. Pham, T. Tran, D. Phung, S. Venkatesh, Predicting healthcare trajectories from medical records: A deep learning approach, *J. Biomed. Inf.*, **69** (2017), 218–229. https://doi.org/10.1016/j.jbi.2017.04.001
10. S. M. Lauritsen, M. E. Kalør, E. L. Kongsgaard, K. M. Lauritsen, M. J. Jørgensen, J. Lange, et al., Early detection of sepsis utilizing deep learning on electronic health record event sequences, *Artif. Intell. Med.*, **104** (2020), 101820. https://doi.org/10.1016/j.artmed.2020.101820
11. M. Aczon, D. Ledbetter, L. Ho, A. Gunny, A. Flynn, J. Williams, et al., Dynamic mortality risk predictions in pediatric critical care using recurrent neural networks, preprint, arXiv:1701.06675.
12. H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, A. Galstyan, Multitask learning and benchmarking with clinical time series data, *Sci. Data*, **6** (2019), 1–18. https://doi.org/10.1038/s41597-019-0103-9
13. S. Hong, S. Lee, J. Lee, W. Cha, K. Kim, Prediction of cardiac arrest in the emergency department based on machine learning and sequential characteristics: model development and retrospective clinical validation study, *JMIR Med. Inf.*, **8** (2020), 1−14. https://doi.org/10.2196/15932
14. Y. J. Lee, K Cho, O. Kwon, H. Park, Y. Lee, J. Kwon, et al., A multicentre validation study of the deep learning-based early warning score for predicting in-hospital cardiac arrest in patients admitted to general wards, *Resuscitation*, **163** (2021), 78−85. https://doi.org/10.1016/j.resuscitation.2021.04.013
15. S. J. Park, K. Cho, O. Kwon, H. Park, Y. Lee, W. H. Shim, et al., Development and validation of a deep-learning-based pediatric early warning system: A single-center study, *Biomed. J.*, **45** (2022), 155−168. https://doi.org/10.1016/j.bj.2021.01.003
16. M. Moor, M. Horn, B. Rieck, D. Roqueiro, K. Borgwardt, Temporal convolutional networks and dynamic time warping can drastically improve the early prediction of sepsis, preprint, arXiv:1902.01659.
17. S. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, preprint, arXiv:1803.01271.
18. Y. Chang, J. Rubin, G. Boverman, S. Vij, A. Rahman, A. Natarajan, et al., A multi-task imputation and classification neural architecture for early prediction of sepsis from multivariate clinical time series, in *2019 Computing in Cardiology (CinC)*, (2019), 1−4. https://doi.org/10.23919/CinC49843.2019.9005751

19. N. Sato, E. Uchino, R. Kojima, S. Hiragi, M. Yanagita, Y. Okuno, Prediction and visualization of acute kidney injury in intensive care unit using one-dimensional convolutional neural networks based on routinely collected data, *Comput. Methods Programs Biomed.*, **206** (2021), 106129. https://doi.org/10.1016/j.cmpb.2021.106129
20. O. Almqvist, A comparative study between algorithms for time series forecasting on customer prediction: An investigation into the performance of ARIMA, RNN, LSTM, TCN and HMM, 2019. Available from: https://www.researchgate.net/publication/333731678
21. G. Montavon, S. Lapuschkin, A. Binder, W. Samek, K. R. Müller, Explaining nonlinear classification decisions with deep taylor decomposition, *Pattern Recognit.*, **65** (2017), 211–222. https://doi.org/10.1016/j.patcog.2016.11.008
22. C. Xiao, E. Choi, J. Sun, Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review, *J. Am. Med. Inf. Assoc.*, **25** (2018), 1419–1428. https://doi.org/10.1093/jamia/ocy068
23. B. Shickel, P. J. Tighe, A. Bihorac, P. Rashidi, Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis, *IEEE J. Biomed. Heal. Inf.*, **22** (2017), 1589–1604. https://doi.org/10.1109/jbhi.2017.2767063
24. W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, K. R. Müller, Explaining deep neural networks and beyond: A review of methods and applications, *Proc. IEEE*, **109** (2021), 247–278. https://doi.org/10.1109/jproc.2021.3060483
25. H. Chefer, S. Gur, L. Wolf, Transformer interpretability beyond attention visualization, in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 782−791. https://doi.org/10.1109/cvpr46437.2021.00084
26. W. J. Nam, S. Gur, J. Choi, L. Wolf, S. W. Lee, Relative attributing propagation: Interpreting the comparative contributions of individual units in deep neural networks, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **34** (2020), 2501–2508. https://doi.org/10.1609/aaai.v34i03.5632
27. S. Gur, A. Ali, L. Wolf, Visualization of Supervised and Self-Supervised Neural Networks via Attribution Guided Factorization, preprint, arXiv:2012.02166.
28. S. M. Lauritsen, M. Kristensen, M. V. Olsen, M. S. Larsen, K. M. Lauritsen, M. J. Jørgensen, et al., Explainable artificial intelligence model to predict acute critical illness from electronic health records, *Nat. Commun.*, **11** (2020), 1–11. https://doi.org/10.1038/s41467-020-17431-x