

**Original Paper**

# Early Prediction of Cardiac Arrest Based on Time-Series Vital Signs Using Deep Learning: Retrospective Study

Yong Li<sup>1</sup>, D; Lei Lv<sup>1</sup>, M; Xia Wang<sup>2</sup>, M

<sup>1</sup>College of Artificial Intelligence and Computer Science, Northwest Normal University, Lanzhou, China

<sup>2</sup>Department of Pharmacy, the People's Hospital of Gansu Province, Lanzhou, China

**Corresponding Author:**

Lei Lv, M  
College of Artificial Intelligence and Computer Science  
Northwest Normal University  
No. 967, Anning East Road  
Lanzhou  
China  
Phone: 86 9317971821  
Email: [lvlei\\_0722@163.com](mailto:lvlei_0722@163.com)

## Abstract

**Background:** Cardiac arrest (CA), characterized by an extremely high mortality rate, remains one of the most pressing global public health challenges. It not only causes a substantial strain on health care systems but also severely impacts individual health outcomes. Clinical evidence demonstrates that early identification of CA significantly reduced the mortality rate. However, the developed CA prediction models exhibit limitations such as low sensitivity and high false alarm rates. Moreover, issues with model generalization remain insufficiently addressed.

**Objective:** The aim of this study was to develop a real-time prediction method based on clinical vital signs, using patient vital sign data from the past 2 hours to predict whether CA would occur within the next 1 hour at 5-minute intervals, thereby enabling timely and accurate prediction of CA events. Additionally, the eICU-CRD dataset was used for external validation to assess the model's generalization capability.

**Methods:** We reviewed and analyzed 4063 patients from the MIMIC-III waveform database, extracting 6 features to develop a deep learning-based CA prediction model named TrGRU. To further enhance performance, statistical features based on a sliding window were also constructed. The TrGRU model was developed using a combination of transformer and gated recurrent unit architectures. The primary evaluation metrics for the model included accuracy, sensitivity, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC), with generalization capability validated using the eICU-CRD dataset.

**Results:** The proposed model yielded an accuracy of 0.904, sensitivity of 0.859, AUROC of 0.957, and AUPRC of 0.949. The results showed that the predictive performance of TrGRU was superior to that of the models reported in previous studies. External validation using the eICU-CRD achieved a sensitivity of 0.813, an AUROC of 0.920, and an AUPRC of 0.848, indicating excellent generalization capability.

**Conclusions:** The proposed model demonstrates high sensitivity and a low false-alarm rate, enabling clinical health care providers to predict CA events in a more timely and accurate manner. The adopted meta-learning approach effectively enhances the model's generalization capability, showcasing its promising clinical application.

(*JMIR Form Res* 2026;10:e78484) doi: [10.2196/78484](https://doi.org/10.2196/78484)

**KEYWORDS**

cardiac arrest; prediction; electric medical record; vital signs; deep learning

## Introduction

Cardiac arrest (CA) is characterized by the abrupt cessation of the heart's blood-pumping function, marked by the absence of

major arterial pulsations and heart sounds. This leads to severe ischemia and hypoxia in vital organs (eg, the brain), ultimately culminating in death. The most common cause of CA is ventricular fibrillation [1]. In the United States, there are

approximately 209,000 cases of in-hospital CA each year, and the global survival rate is less than 25%. Previous studies have demonstrated that early recognition of CA can improve the survival rate within the first hour by approximately 29% and before hospital discharge by 19% [2]. Therefore, it is essential to develop a model that can reliably predict the occurrence of CA events.

In clinical practice, traditional assessment methods based on scoring, such as the Simplified Acute Physiology Score II, Sequential Organ Failure Assessment, and Modified Early Warning Score, have long served as crucial reference tools for clinicians to identify the risk of clinical deterioration and initiate timely interventions [3-5]. However, such scoring systems generally suffer from the dual limitations of low sensitivity and high false alarm rates. In recent years, machine learning has been applied extensively in health care data analysis, showing significantly better predictive performance than traditional scoring systems. Nevertheless, machine learning still faces challenges in analyzing high-dimensional and complex time-series data, which makes these approaches inadequate to meet the demands of current clinical practice.

Kim et al [6] developed a CA prediction model using the TabNet classifier to predict CA events. Their model achieved an area under the receiver operating characteristic curve (AUROC) of 0.79 on the MIMIC-IV dataset. Wu et al [7] developed a CA prediction model using extreme gradient boosting (XGBoost) based on 20 variables, including vital signs, laboratory results, and electrocardiogram reports. The model accuracy was 0.889, and the AUROC was 0.958. Similarly, Yijing et al [8] extracted vital sign data from the MIMIC-III database and used XGBoost to develop a model that achieved an accuracy of 0.96 and an AUROC of 0.94. An early warning model based on a recurrent neural network was proposed by Kwon et al [9], which yielded an area under the precision-recall curve (AUPRC) of 0.04 and an AUROC of 0.85.

Current methods based on deep learning and machine learning have demonstrated good performance in predicting CA events. However, they still suffer from low sensitivity and high false alarm rates as well as the inability to make real-time predictions. In particular, the method proposed by Lee et al [10] exhibited significant performance fluctuations when predicting CA events within 24 hours, a limitation that has also been demonstrated in previous studies. These shortcomings will lead to a substantial waste of medical resources and an increase in operational costs. As a result, current CA prediction methods face significant challenges in practical applications.

In this study, we propose TrGRU, a deep learning-based model that accurately predicts CA events within an hour using only 6 limited clinical vital signs without relying on patient laboratory test results. Meanwhile, we adopt a meta-learning framework, which significantly improves the adaptability of the model across different datasets and clinical settings [11,12].

## Methods

### Dataset

In this study, we used 2 databases: the MIMIC-III database [13] and eICU-CRD [14]. The MIMIC-III database, developed by the Massachusetts Institute of Technology Laboratory for Computational Physiology, contains data from 53,432 adult patients and 8100 neonatal patients admitted between 2001 and 2012. The database includes demographic data, vital signs, medications, laboratory measurements, physician orders, procedure codes, diagnosis codes, imaging reports, hospitalization duration, and survival data, among others.

The eICU-CRD is a multicenter intensive care unit (ICU) database that covers data from more than 200,000 ICU admissions across 208 hospitals in the United States between 2014 and 2015. The data were collected through Philips eICU program, which includes vital sign measurements, nursing documentation, severity of illness scores, diagnostic information, treatment details, and more.

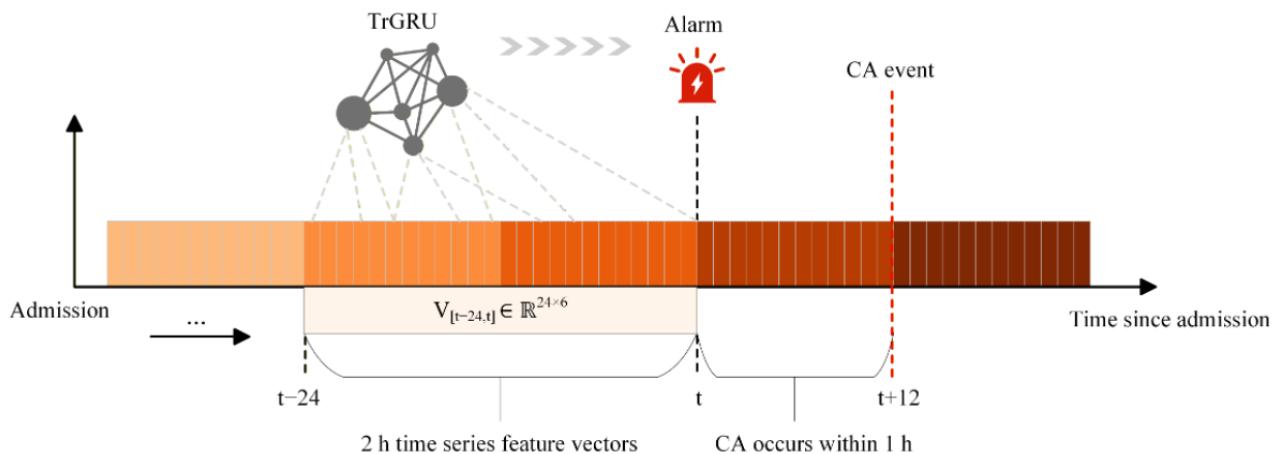
We randomly allocated 70% (2844/4063) of the dataset for the training set to develop the model, with 10% (406/4063) as the validation set to adjust and determine the hyperparameters of the model, and the remaining 20% (813/4063) as the test set to evaluate the model performance; this portion was not used in the model training.

### Ethical Considerations

This study used publicly available critical care databases, including the MIMIC-III and the eICU-CRD. Both databases received institutional review board approval from the hospitals that originally contributed the data, and informed consent was obtained at the time of data collection. All patient records were deidentified in accordance with the Health Insurance Portability and Accountability Act (HIPAA). Access to the databases was granted through the PhysioNet credentialing process. Therefore, according to institutional policies regarding secondary research using publicly deidentified data, additional ethics approval from our institution was not required.

### Problem Definition

As shown in Figure 1, let the given set of patients be  $P = \{p_1, p_2, \dots, p_n\}$  feature vector  $V_t = [hr, rr, sbp, dbp, map, spo2]$ .  $V_t$  represents the vital sign monitoring data of patient  $p$  at time point  $t$ , including “heart rate” and “systolic blood pressure,” among others. These data were continuously collected during the hospitalization period  $T$ ,  $T = \{t_1, t_2, \dots, t_j\}$  and  $t_i - t_{i-j} = 5$  minutes. Then,  $V = \bigcup_{i=1}^j V_t$  represents the multivariate time-series feature vectors of the patient at all time points during hospitalization. Therefore, the CA prediction task can be formally described as follows: given the multivariate time-series feature vector  $V$  for patient  $p$  in the hospitalization interval  $T$ , the goal of the prediction task is, for any time point  $t_i \in T$ , to use the 24 multivariate time-series feature vectors from the time window  $(t-24 \text{ and } t)$  to predict the probability of a CA event occurring within the time window  $(t \text{ and } t+12)$  using the TrGRU model.

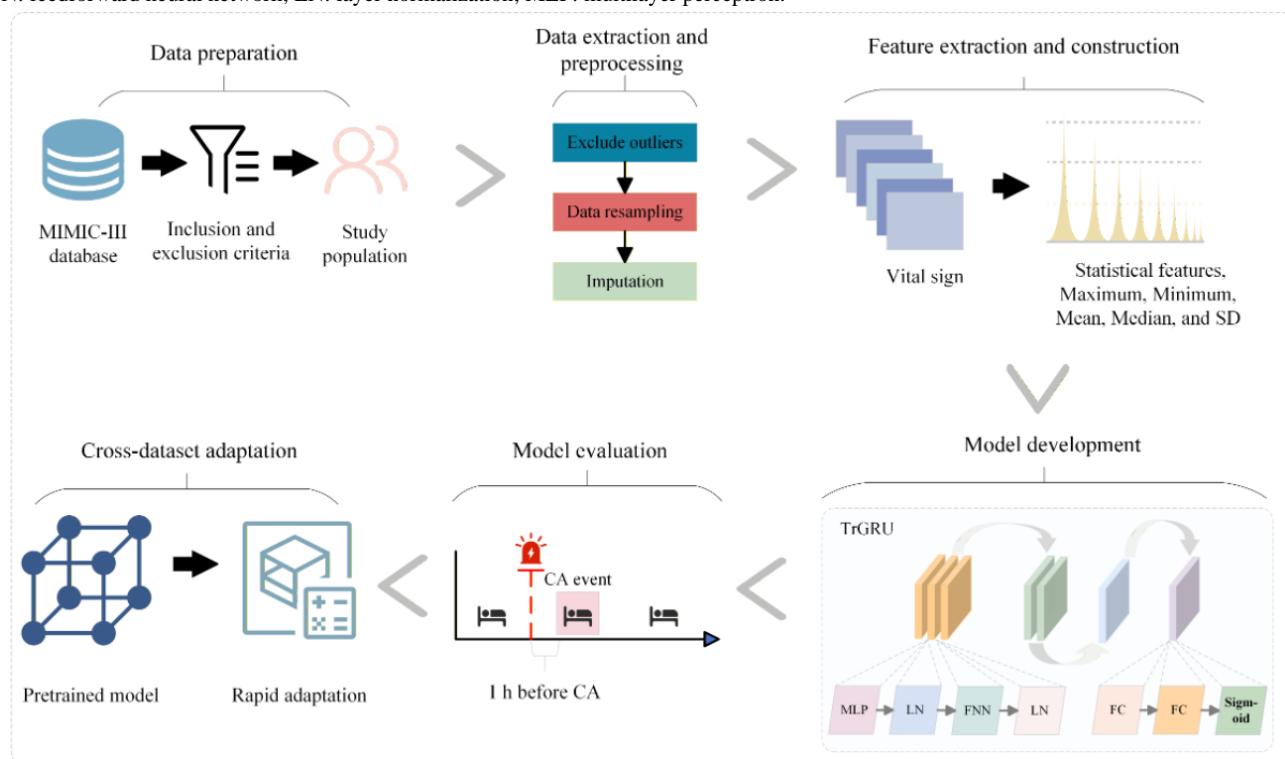
**Figure 1.** Cardiac arrest (CA) event monitoring process for the proposed TrGRU model.

## Development Process

### Review

As shown in [Figure 2](#), the development process of the TrGRU model consisted of 6 steps: data preparation, data extraction and

**Figure 2.** Development process of the TrGRU model. The development process of the TrGRU model consists of 6 steps. FC: fully connected layer; FNN: feedforward neural network; LN: layer normalization; MLP: multilayer perceptron.



### Data Preparation

We extracted data from the MIMIC-III databases and eICU-CRD to construct a study cohort that met the inclusion and exclusion criteria. The patients in these 2 databases differ. The MIMIC-III database includes patients admitted to the ICU as well as some hospitalized patients not admitted to the ICU, while the eICU-CRD database contains only patients admitted to the ICU. The number of CA events for each patient also varies between the 2 databases. While each patient in the eICU-CRD frequently records several CA events, each patient in the

and preprocessing, feature extraction and construction, model development, model evaluation, and cross-dataset adaptation.

MIMIC-III database typically only records one CA event. To increase the sample size of CA events, we treated multiple CA events experienced by a single patient as independent samples for model development.

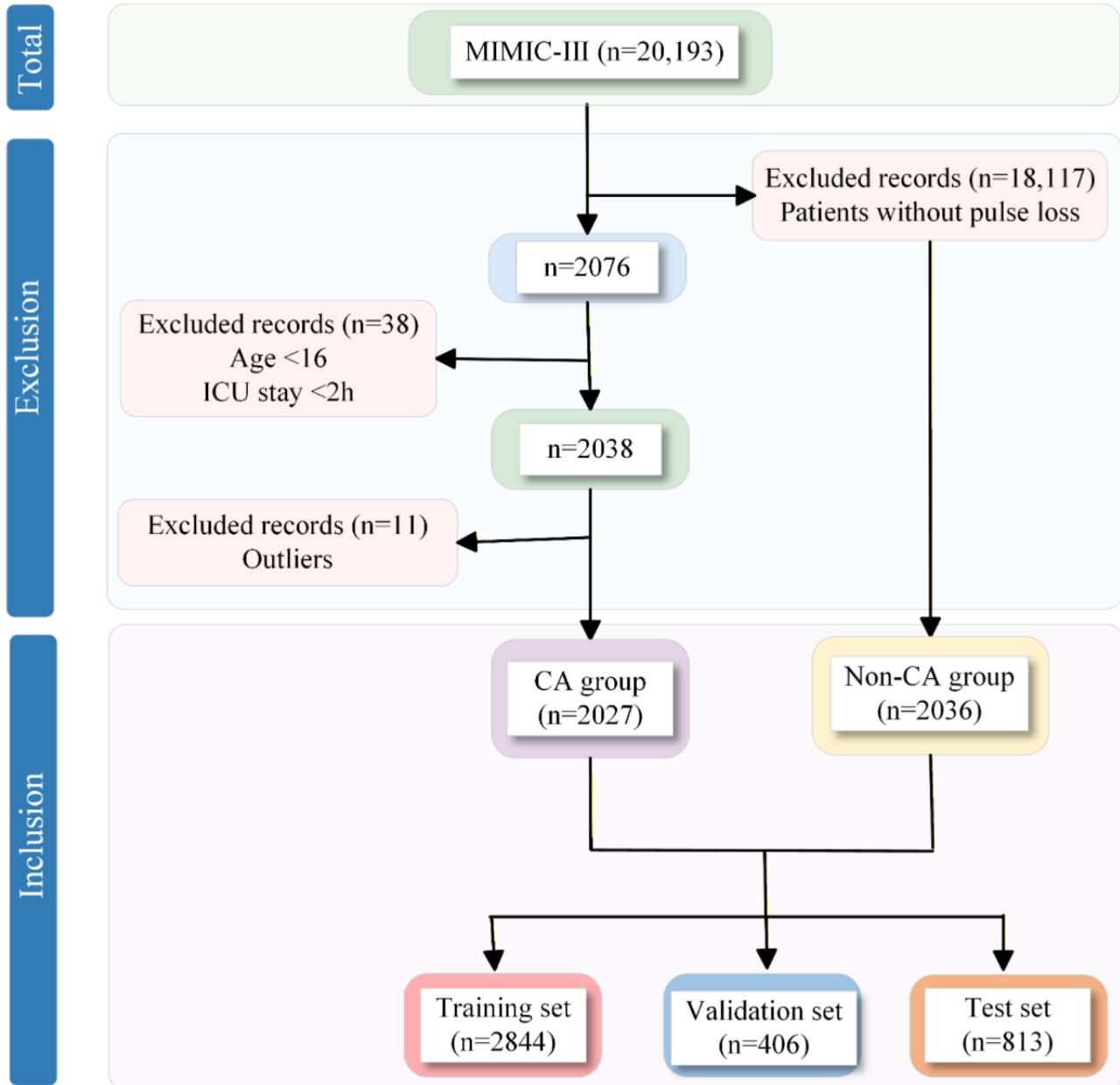
A series of criteria were applied during the inclusion and exclusion process of the MIMIC-III database to select the study cohort. As shown in [Figure 3](#), 20,193 patients were included in this study, with 38 (0.19%) patients younger than 16 years and those who experienced CA within 2 hours of admission being excluded. In addition, 11 (0.05%) patients with outliers were

excluded. CA is defined as the loss of a detectable pulse during attempts at resuscitation, and patients in this category were included in the CA group. Patients in the non-CA group were randomly selected from the remaining patients who did not experience a loss of pulse. Ultimately, the study population

consisted of 4063 (20.12%) patients, with 2027 (49.89%) in the CA group and 2036 (50.11%) in the non-CA group of whom were not.

The inclusion and exclusion criteria applied to the eICU-CRD were the same as those used for the MIMIC-III database.

**Figure 3.** Patient inclusion and exclusion flow diagram for the MIMIC-III. CA: cardiac arrest; ICU: intensive care unit.



### Data Extraction and Preprocessing

In this study, patient vital sign data were extracted from the MIMIC-III database, including heart rate (HR), respiratory rate, systolic blood pressure, diastolic blood pressure, mean arterial pressure, and oxygen saturation ( $\text{SpO}_2$ ). Due to the potential for data errors caused by equipment malfunction, the acceptable range for each vital sign value was determined based on the advice of medical experts. Data falling outside these ranges were considered outliers and excluded. Subsequently, the vital signs time-series data with a 1-minute sampling frequency were resampled to a 5-minute frequency. For time points without sampled values, these were marked as missing. We used median

imputation to handle these missing values [15]. This method fills in missing values by identifying the median of the data, which is less affected by extreme values and thus provides a relatively stable substitute.

Additionally, because each feature is measured on a different scale, we applied minimum-maximum normalization to standardize them. Minimum-maximum normalization is a linear transformation method that maps data values to a fixed range. In this study, the values were scaled to the range (0,1). This process is achieved by scaling the minimum and maximum values of the data, ensuring that the transformed values are independent of the scale and range of the original data.

Although the number of patients included in the CA and non-CA groups was nearly equal, the samples input into the model were generated using a 2-hour sliding time window, and CA events occurred far less frequently than nonevents; that is, the number of positive samples was substantially lower than that of negative samples, resulting in class imbalance at the window level. Therefore, we mitigated the impact of class imbalance on model training by increasing the number of positive samples through oversampling and reducing the number of negative samples through undersampling.

### Feature Extraction and Construction

We used 6 preprocessed vital signs as model features: HR, respiratory rate, systolic blood pressure, diastolic blood pressure, mean arterial pressure, and  $\text{SpO}_2$ . These vital signs are widely used in clinical practice, and abnormal fluctuations in these measures are associated with the occurrence of CA. Previous studies also demonstrated that models developed based on these vital signs performed well in predicting CA events, making them suitable as model features [16,17].

In addition, we also created statistical features based on a sliding window. A fixed-length sliding window of 2 hours was applied, with a 5-minute time step, to segment each vital sign. The mean, median, minimum, maximum, and SD of each feature were then

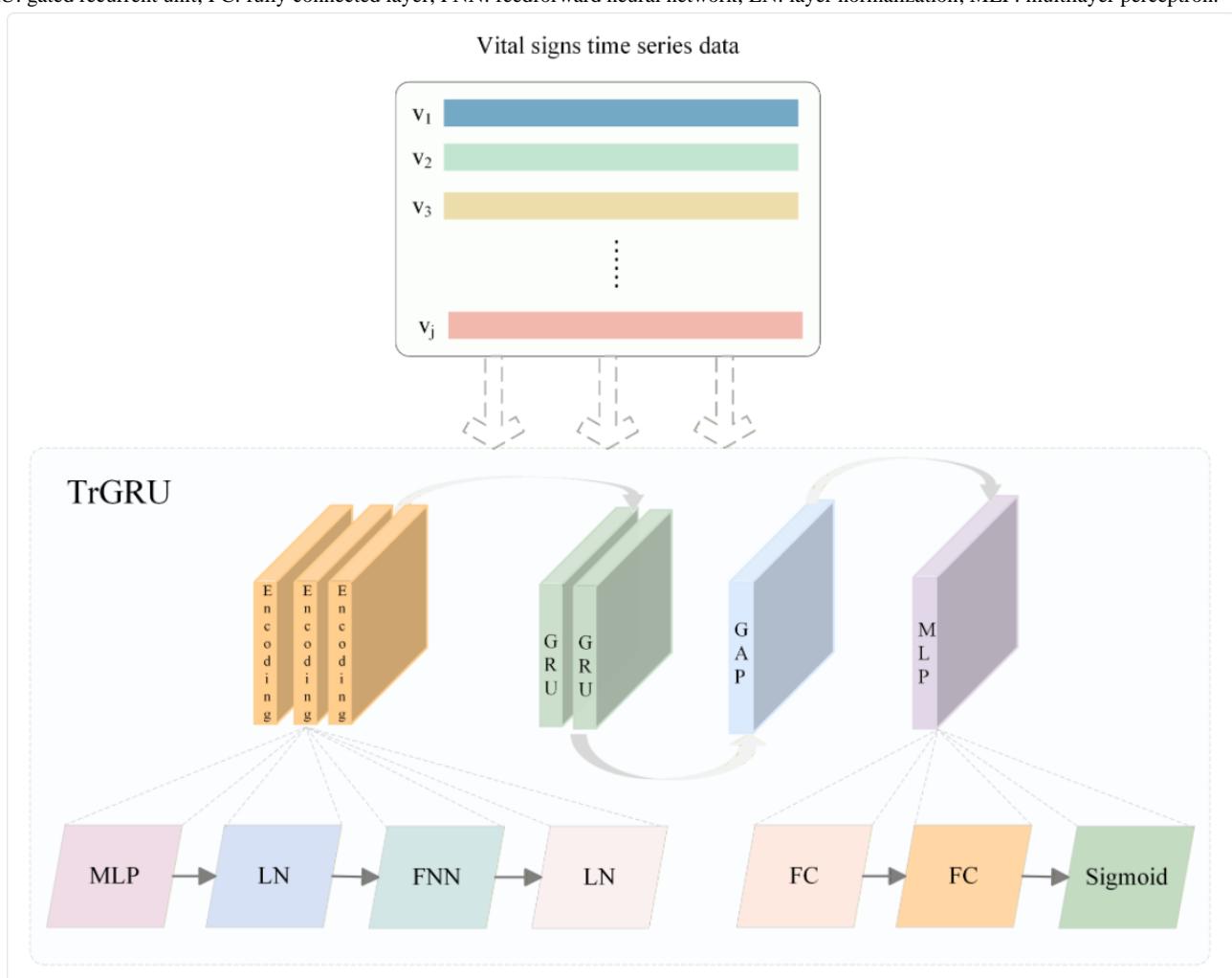
calculated for the time-series segments of the vital sign data within each window.

### Model Development

We developed the TrGRU model based on deep learning methods, transformer [18] and gated recurrent unit (GRU) [19], using 2-hour time window vital sign data to predict CA events within the next 1 hour. By inputting the 2-hour data into the model, the risk score for a patient experiencing CA within the next 1 hour was evaluated. When the predicted score exceeded the risk threshold, it indicated a higher likelihood of the patient experiencing CA within 1 hour, and the patient was labeled as an event.

As shown in [Figure 4](#), the TrGRU is a hybrid model that combines a transformer and a GRU, designed to handle time-series data. Typically, the transformer is composed of an encoder and a decoder. In the TrGRU architecture, 3 encoder layers were stacked first, followed by 2 GRU layers. After the GRU layers, global average pooling is applied to compress the dimensionality of the time-series data, which pools the output feature vector of each sequence into a fixed-length vector. The pooled output was further processed by a multilayer perceptron head, in which the decoder was replaced with fully connected layers, as decoding was no longer necessary.

**Figure 4.** Architecture diagram of the TrGRU model.  $V_1, V_2, V_3$ , and  $V_j$  are multivariate time-series feature vectors. GAP: global average pooling; GRU: gated recurrent unit; FC: fully connected layer; FNN: feedforward neural network; LN: layer normalization; MLP: multilayer perceptron.



## Model Evaluation

### Baseline Models

To evaluate the performance of the proposed model, it was compared with 4 baseline models: random forest, logistic regression, XGBoost, and light gradient boosting method (LGBM).

### Evaluation Metrics

We used accuracy, sensitivity, specificity, AUROC, AUPRC, and the false alarm rate to evaluate the performance of the model. AUROC is a metric that measures the performance of a binary classification model across different thresholds. The closer the AUROC value is to 1, the better the classification ability of the model. Sensitivity is also an important performance metric, representing the ability of the model to correctly identify actual positive samples (CA events). Developing a highly sensitive model is essential for the CA prediction task because a model with low sensitivity may miss CA events, which could lead to severe consequences for patients. Therefore, the goal of this study was to develop a prediction model with high sensitivity and a low false alarm rate. The false alarm rate is defined as the proportion of incorrectly detected CA events among all alarms. The calculation of the false alarm rate is as follows:

$$\text{FAR} = 1 - (\text{N}_{\text{True}} / \text{N}_{\text{All}})$$

where FAR is the false alarm rate,  $\text{N}_{\text{True}}$  is the number of correct alarms, and  $\text{N}_{\text{All}}$  is the total number of alarms.

### Cross-Dataset Adaptation

We used meta-learning combined with fast adaptation to solve the problem of model adaptation across different datasets, aiming to equip the model with the ability to learn new tasks quickly and adapt to unseen tasks with a small number of samples [20,21]. Specifically, the model was first pretrained on the MIMIC-III dataset to acquire a general learning strategy, thus achieving a good initial state that allows for rapid adjustment when faced with new tasks. Once the model was pretrained, it underwent fast adaptation on the eICU-CRD dataset, where parameters were quickly adjusted using a small amount of data to improve performance on the new dataset. With this approach, the issue of poor generalization capability of existing models was solved, allowing for better adaptation to various clinical settings.

## Experimental Setup

### Overview

In addition to model evaluation, comparison with baseline models, and cross-dataset adaptation experiments, 3 other sets of experiments were conducted: evaluation of time-phased prediction performance, the impact of feature sets on prediction performance, and the trends of feature changes over time. The subsequent sections will introduce these 3 sets of experiments.

### Performance Evaluation of Time-Phased Prediction

This experiment evaluated the ability of the model to predict CA 30, 20, and 10 minutes in advance. The primary evaluation metric was sensitivity, which assessed the ability of the model to identify CA events.

### Impact of Feature Set on Prediction Performance

We constructed 3 types of feature sets: raw features based on vital signs, statistical features, and a combination of raw and statistical features. Various feature set experiments were conducted to determine the impact of each feature set on the performance of the model.

### Trends of Feature Changes Over Time

We analyzed the average trends of different vital sign values in the CA and non-CA groups over the 30 minutes before the occurrence of CA and visualized these trends.

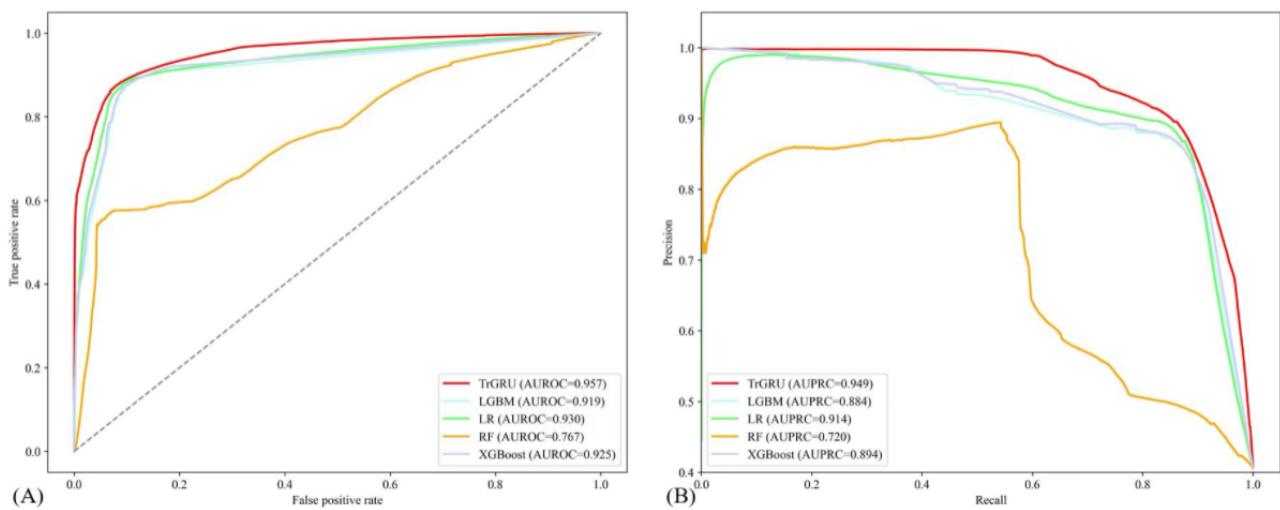
## Results

### Model Performance Evaluation

This section presents the evaluation results of the model prediction performance, which were evaluated using accuracy, sensitivity, specificity, AUROC, AUPRC, and false alarm rate. Additionally, the performance of the proposed model was compared with that of the baseline models.

In the MIMIC-III dataset, several models used in existing studies were compared with the proposed model to evaluate its performance in predicting CA events. As shown in [Figure 5](#), the proposed model achieved an AUROC of 0.957 and an AUPRC of 0.949, significantly outperforming the baseline models. In addition, other performance metrics were also compared to fully assess the validity of the model in relevant clinical contexts. As shown in [Table 1](#), the proposed model achieved an accuracy of 0.904, a sensitivity of 0.859, a specificity of 0.933, and a false alarm rate of 0.067, all of which outperform the baseline models, yielding the best performance among these.

**Figure 5.** (A) The receiver operating characteristic (ROC) curve and (B) the precision-recall curve on the test set. AUROC: area under the receiver operating characteristic curve; AUPRC: area under the precision-recall curve; LGBM: light gradient boosting method; XGBoost: extreme gradient boosting.



**Table 1.** Performance comparison of the proposed TrGRU model with baseline models.

Model	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	False alarm rate (95% CI)	Area under the receiver operating characteristic curve (95% CI)	Area under the precision-recall curve (95% CI)
Logistic regression	0.895 (0.895-0.897)	0.841 (0.851-0.854)	0.931 (0.925-0.927)	0.069 (0.073-0.075)	0.930 (0.930-0.931)	0.914 (0.913-0.915)
Extreme Gradient Boosting	0.890 (0.889-0.891)	0.852 (0.851-0.854)	0.916 (0.915-0.916)	0.084 (0.084-0.085)	0.925 (0.924-0.925)	0.894 (0.893-0.895)
Light gradient boosting method	0.869 (0.868-0.869)	0.775 (0.773-0.776)	0.933 (0.932-0.934)	0.067 (0.066-0.068)	0.919 (0.918-0.919)	0.884 (0.883-0.885)
Random forest	0.663 (0.662-0.664)	0.708 (0.706-0.709)	0.633 (0.631-0.634)	0.367 (0.366-0.369)	0.767 (0.766-0.768)	0.720 (0.717-0.722)
TrGRU	0.904 (0.903-0.904)	0.859 (0.858-0.861)	0.933 (0.933-0.935)	0.067 (0.065-0.067)	0.957 (0.956-0.957)	0.949 (0.949-0.950)

## External Validation

We used a meta-learning and rapid adaptation-based approach to perform the cross-dataset CA prediction task. On the MIMIC-III dataset, a model was pretrained through multitask meta-learning. After pretraining, the lower-layer parameters of the model were frozen on the eICU-CRD, and only the last 2 fully connected layers were fine-tuned. Domain adaptation was achieved through adaptive training. The experimental results showed that the model achieved a sensitivity of 0.813, an AUROC of 0.920, and an AUPRC of 0.848 on the independent

eICU-CRD, demonstrating significantly better cross-dataset adaptation ability compared to current models. These results indicated that the framework effectively used knowledge from the source domain to enhance the few-shot learning performance in the target domain.

## Performance Evaluation of Time-Phased Prediction

We also evaluated the ability of the model to predict CA 30, 20, and 10 minutes in advance, using sensitivity, specificity, AUROC, AUPRC, and the false alarm rate as evaluation metrics, as shown in [Table 2](#).

**Table 2.** Predictive performance of the TrGRU model across different time windows before cardiac arrest (CA).

Before CA	Sensitivity	Specificity	Area under the receiver operating characteristic curve	Area under the precision-recall curve	False alarm rate
30 min	0.906	0.926	0.967	0.955	0.073
20 min	0.926	0.923	0.971	0.958	0.077
10 min	0.948	0.919	0.977	0.963	0.081

The proposed model could identify 90.6% (737/813) of the patients experiencing CA 30 minutes in advance, 92.6% (753/813) of the patients 20 minutes in advance, and 94.8% (771/813) of the patients 10 minutes in advance, with a false alarm rate of 8% or lower within 30 minutes. Compared to other studies, the model proposed in this study exhibited higher sensitivity and a lower false alarm rate.

**Table 3.** Impact of different feature sets on model prediction performance.

Feature sets	Accuracy	Sensitivity	Specificity	Precision	$F_1$ -score	Area under the receiver-operating characteristic curve	Area under the precision-recall curve
Raw features	0.904	0.859	0.933	0.898	0.878	0.957	0.949
Statistical features	0.918	0.894	0.934	0.903	0.898	0.980	0.972
Combination features	0.918	0.892	0.935	0.904	0.898	0.981	0.973

Regarding accuracy, sensitivity, AUROC, AUPRC, and  $F_1$ -score, the model performed better when statistical features were used than when raw features were used. When a combination of statistical and raw features was fed into the model, its predictive performance was also superior to that obtained using only raw features. However, its performance showed almost no difference compared to using only statistical features. Consequently, it can be inferred that statistical features enhanced the prediction performance of the model.

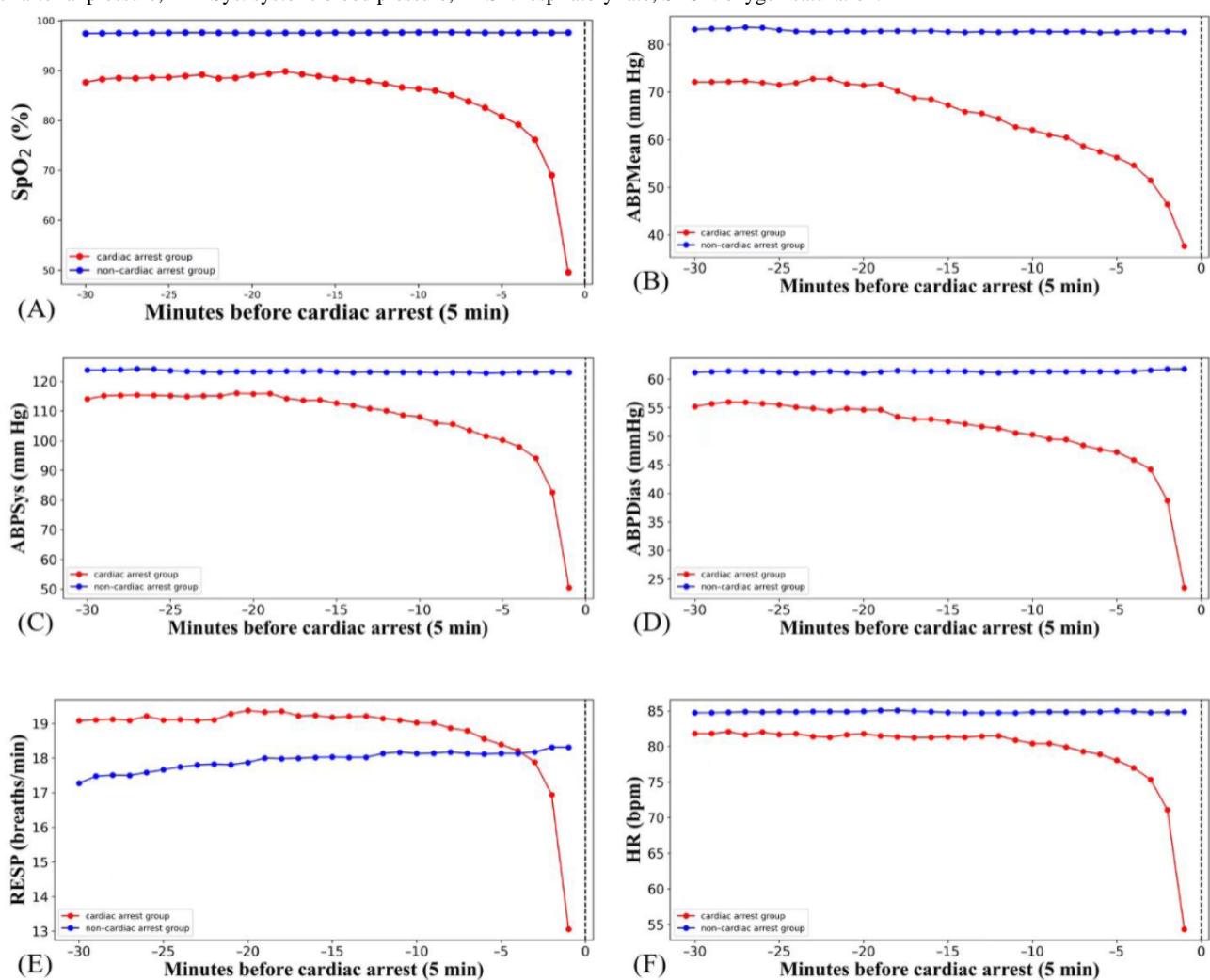
### Impact of Feature Set on Prediction Performance

We constructed 3 types of feature sets: raw features based on vital signs, statistical features, and a combination of raw features and statistical features. The 3 types of feature sets were input into the model for training to determine the impact of different feature sets on the model prediction performance. The results are shown in Table 3.

### Trends of Feature Changes Over Time

Vital sign patterns over time for patients in the CA group and the non-CA group are shown in Figure 6. Compared to the non-CA group, the CA group exhibited lower HR, SpO<sub>2</sub>, and blood pressure values before the event, while respiratory rate was slightly higher. Vital signs began to decline 25 minutes before CA. The decline became more noticeable within the 10 minutes preceding the event. Among all the vital signs, blood pressure and SpO<sub>2</sub> showed an earlier decline. Compared to the CA group, the vital signs of the non-CA group remained relatively stable.

**Figure 6.** Average trends of vital signs over time in the cardiac arrest and non–cardiac arrest groups. ABPDias: diastolic blood pressure; ABPMean: mean arterial pressure; ABPSys: systolic blood pressure; RESP: respiratory rate; SPO2: oxygen saturation.



## Discussion

### Principal Findings

A deep learning model with real-time capability, high sensitivity, and a low false alarm rate was developed using only 6 vital signs based on the MIMIC-III database. The model relied on clinically accessible indicators to determine whether a patient would experience CA within the next hour at 5-minute intervals. It enabled real-time and continuous monitoring by overcoming the limitations of traditional prediction models that rely on a large number of features or features not commonly used in

hospitals. The outcomes demonstrated that the model achieved excellent prediction performance. The TrGRU model developed in this study outperformed logistic regression, LGBM, and random forest across all metrics. Notably, compared to other models, TrGRU achieved both high sensitivity and a low false alarm rate. Table 4 presents a comparison between this study and other studies. Layeghian et al [22] developed a prediction model using ensemble learning methods based on multivariate features, vital signs, and clinical latent features, achieving an AUROC of 0.82 on the MIMIC-III dataset. However, their model was not designed for real-time monitoring.

**Table 4.** Comparison with previous studies.

Study	Database	Features	Model	Real time	Cross-validation	Area under the receiver operating characteristic curve
Kim et al [6]	MIMIC-IV	Vital signs, statistical Features, and Gini index	TabNet classifier	Yes	Yes	0.86
Kwon et al [9]	Clinical database	Vital signs	Recurrent neural network	Yes	Yes	0.85
Yijing et al [8]	MIMIC-III	Vital signs	Extreme Gradient Boosting	Yes	No	0.94
Layeghian et al [22]	MIMIC-III	Multivariate, vital signs, and clinical latent	Ensemble learning	No	No	0.82
Our work	MIMIC-III	Vital signs	TrGRU	Yes	Yes	0.96

The experiments investigating the impact of different feature sets on model prediction performance demonstrated that statistical features outperformed raw features in accuracy, sensitivity, AUROC, and AUPRC. When a combination of raw and statistical features was used, performance was also superior to that obtained using raw features alone but comparable to that obtained using only statistical features. Therefore, statistical features played a crucial role in predicting CA events.

Class imbalance is a critical issue that affects model training performance. When the data are severely imbalanced, models often perform poorly in predicting minority classes (ie, they tend to favor predicting negative samples). This situation is particularly common in medical datasets, as in the context of this study, the occurrence of CA events was much rarer than in normal conditions. To address this problem, we used upsampling to replicate positive samples and downsampling to reduce negative samples, thereby balancing the ratio of positive and negative class samples. This solution effectively improved the sensitivity of the model (sensitivity increased from 80.6% before sampling to 85.9% after sampling).

The TrGRU can identify 85.9% (689/813) of the patients 1 hour before a CA event, 90.6% (737/813) of the patients 30 minutes before, and 92.6% (753/813) and 94.8% (771/813) of the patients 20 and 10 minutes before, respectively. These results imply that even when it is too late to prevent CA, medical staff still have a certain amount of time to intervene, thereby improving patient survival rates. This is because the faster cardiopulmonary resuscitation is performed after CA, the greater the patient's chance of survival, with the survival rate decreasing by 10% per minute before cardiopulmonary resuscitation is initiated. The model proposed by Yijing et al [8] identified 80% of the patients 25 minutes before CA and 93% of the patients 10 minutes before the event. The model proposed by Kwon et al [9] could also only identify 78% of the patients 30 minutes before CA. Therefore, the model proposed in this study demonstrated excellent performance in predicting CA events in advance, outperforming some existing studies.

Due to the “black box” nature of deep learning, it is challenging to establish relationships between real-time prediction results and features. In contrast, TrGRU provided a certain level of interpretability. We analyzed patterns of feature values over

time by comparing the changes in vital signs between the CA group and the non-CA group before the occurrence of CA. This provided health care professionals with an objective basis for assessing physiological conditions. Clinical studies have shown that synergistic abnormal changes in vital signs (such as a sharp increase in respiratory rate combined with abnormal fluctuations in body temperature) often exhibit significant characteristic patterns before the occurrence of CA [23-25]. These patterns provide crucial evidence for clinical prediction and the determination of intervention windows.

To improve model generalizability, we used a meta-learning approach to address the issue of the model's inability to adapt to different datasets. Using the MIMIC-III dataset as the pretraining dataset and the eICU-CRD dataset as the target for rapid adaptation, the pretrained model requires only a small number of samples to quickly adjust and adapt to the new dataset. The results demonstrated that the method proposed in this study significantly improved the generalization of the model's capability and achieved excellent performance. Kim et al [6] trained their model on the MIMIC-IV dataset and subsequently conducted cross-dataset testing on the eICU-CRD dataset, achieving an AUROC of 0.80. Kwon et al [9] also performed cross-dataset validation on their proposed model, ultimately achieving an AUROC of 0.837 and an AUPRC of 0.239.

## Contributions

This study primarily made 4 contributions. First, compared with existing studies, we developed a highly sensitive CA prediction model. Sensitivity exceeded 90% in predicting CA events within 30 minutes, which could allow health care professionals to intervene earlier and improve patient survival rates. Second, we used a meta-learning approach to enhance the generalizability of the model, enabling it to adapt to different datasets rapidly. To the best of our knowledge, this was the first study to use meta-learning to address the adaptability of medical prediction models across different datasets. Third, the model in this study significantly reduced the false alarm rate, which could effectively help avoid unnecessary interventions or treatments, thereby improving the efficient use of medical resources and reducing health care costs. Additionally, reducing false alarms helps prevent “alarm fatigue” [26], ensuring that health care

professionals remain highly alert and responsive to each alarm. Fourth, we developed a prediction model using fewer variables, overcoming limitations of existing models that rely on variables not commonly used in hospitals. This approach offers the advantage of being applicable to various hospital settings.

### Limitations

Our work mainly involves 3 limitations. First, due to the “black box” nature of deep learning, it is unable to establish relationships between prediction results and input data, and the interpretability of results is a critical factor for clinicians in making medical decisions; therefore, deep learning-based medical decision support still faces practical challenges. In recent years, the interpretability of deep learning has been explored in research [27-31], which is a key focus for our next steps. Second, CA may be caused by a variety of diseases, and in this study, we did not consider the heterogeneity between patients with different diseases [32]. Therefore, the generalizability of TrGRU cannot be guaranteed. Accordingly, it is necessary to expand the dataset to cover various types of diseases to further enhance the generalizability and robustness of the model. Third, we did not conduct feature selection.

### Funding

This research was partially supported by the grants from the Natural Science Foundation of China (grant 62163033), the Soft Science Special Project of Gansu Basic Research Plan (grant 23JRZA397), and the 2026 Gansu Industry Support Plan Project for University (grant 2026CYZC-011).

### Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

### Authors' Contributions

YL and LL are co-first authors who jointly designed the research; LL performed experiments and wrote the manuscript. YL supervised the project and revised the manuscript. XW contributed clinical expertise and guidance.

### Conflicts of Interest

None declared.

### References

1. Allencherril J, Lee PY, Khan K, Loya A, Pally A. Etiologies of in-hospital cardiac arrest: a systematic review and meta-analysis. *Resuscitation*. Jun 2022;175:88-95. [doi: [10.1016/j.resuscitation.2022.03.005](https://doi.org/10.1016/j.resuscitation.2022.03.005)] [Medline: [35278525](https://pubmed.ncbi.nlm.nih.gov/35278525/)]
2. Bergum D, Haugen BO, Nordseth T, Mjølstad OC, Skogvoll E. Recognizing the causes of in-hospital cardiac arrest--a survival benefit. *Resuscitation*. Dec 2015;97:91-96. [[FREE Full text](#)] [doi: [10.1016/j.resuscitation.2015.09.395](https://doi.org/10.1016/j.resuscitation.2015.09.395)] [Medline: [26449872](https://pubmed.ncbi.nlm.nih.gov/26449872/)]
3. Clemency BM, Murk W, Moore A, Brown LH. The EMS Modified Early Warning Score (EMEWS): a simple count of vital signs as a predictor of out-of-hospital cardiac arrests. *Prehosp Emerg Care*. Apr 13, 2022;26(3):391-399. [doi: [10.1080/10903127.2021.1908464](https://doi.org/10.1080/10903127.2021.1908464)] [Medline: [33794729](https://pubmed.ncbi.nlm.nih.gov/33794729/)]
4. Spångfors M, Molt M, Samuelson K. In-hospital cardiac arrest and preceding National Early Warning Score (NEWS): a retrospective case-control study. *Clin Med (Lond)*. Jan 2020;20(1):55-60. [[FREE Full text](#)] [doi: [10.7861/clinmed.2019-0137](https://doi.org/10.7861/clinmed.2019-0137)] [Medline: [31941734](https://pubmed.ncbi.nlm.nih.gov/31941734/)]
5. Aegerter P, Boumendil A, Retbi A, Minvielle E, Dervaux B, Guidet B. SAPS II revisited. *Intensive Care Med*. Mar 28, 2005;31(3):416-423. [doi: [10.1007/s00134-005-2557-9](https://doi.org/10.1007/s00134-005-2557-9)] [Medline: [15678308](https://pubmed.ncbi.nlm.nih.gov/15678308/)]
6. Kim YK, Seo W, Lee SJ, Koo JH, Kim GC, Song HS, et al. Early prediction of cardiac arrest in the intensive care unit using explainable machine learning: retrospective study. *J Med Internet Res*. Sep 17, 2024;26:e62890. [[FREE Full text](#)] [doi: [10.2196/62890](https://doi.org/10.2196/62890)] [Medline: [39288404](https://pubmed.ncbi.nlm.nih.gov/39288404/)]

7. Wu TT, Lin XQ, Mu Y, Li H, Guo YS. Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes. *Clin Cardiol*. Mar 14, 2021;44(3):349-356. [[FREE Full text](#)] [doi: [10.1002/clc.23541](https://doi.org/10.1002/clc.23541)] [Medline: [33586214](#)]
8. Yijing L, Wenyu Y, Kang Y, Shengyu Z, Xianliang H, Xingliang J, et al. Prediction of cardiac arrest in critically ill patients based on bedside vital signs monitoring. *Comput Methods Programs Biomed*. Feb 2022;214:106568. [doi: [10.1016/j.cmpb.2021.106568](https://doi.org/10.1016/j.cmpb.2021.106568)] [Medline: [34883382](#)]
9. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *J Am Heart Assoc*. Jun 26, 2018;7(13):e35. [[FREE Full text](#)] [doi: [10.1161/JAHA.118.008678](https://doi.org/10.1161/JAHA.118.008678)] [Medline: [29945914](#)]
10. Lee H, Yang H, Ryu HG, Jung C, Cho YJ, Yoon SB, et al. Real-time machine learning model to predict in-hospital cardiac arrest using heart rate variability in ICU. *NPJ Digit Med*. Nov 23, 2023;6(1):215. [[FREE Full text](#)] [doi: [10.1038/s41746-023-00960-2](https://doi.org/10.1038/s41746-023-00960-2)] [Medline: [37993540](#)]
11. Vettoruzzo A, Bouguelia M, Vanschoren J, Rögnvaldsson T, Santosh K. Advances and challenges in meta-learning: a technical review. *IEEE Trans Pattern Anal Mach Intell*. Jul 2024;46(7):4763-4779. [doi: [10.1109/tpami.2024.3357847](https://doi.org/10.1109/tpami.2024.3357847)]
12. Gharoun H, Momenifar F, Chen F, Gandomi AH. Meta-learning approaches for few-shot learning: a survey of recent advances. *ACM Comput. Surv.* Jul 25, 2024;56(12):1-41. [doi: [10.1145/3659943](https://doi.org/10.1145/3659943)]
13. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. May 24, 2016;3:160035. [[FREE Full text](#)] [doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35)] [Medline: [27219127](#)]
14. Pollard TJ, Johnson AE, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data*. Sep 11, 2018;5:180178. [[FREE Full text](#)] [doi: [10.1038/sdata.2018.178](https://doi.org/10.1038/sdata.2018.178)] [Medline: [30204154](#)]
15. Berkelmans GF, Read SH, Gudbjörnsdóttir S, Wild SH, Franzen S, van der Graaf Y, et al. Population median imputation was noninferior to complex approaches for imputing missing values in cardiovascular prediction models in clinical practice. *J Clin Epidemiol*. May 2022;145:70-80. [[FREE Full text](#)] [doi: [10.1016/j.jclinepi.2022.01.011](https://doi.org/10.1016/j.jclinepi.2022.01.011)] [Medline: [35066115](#)]
16. Considine J, Casey P, Omonaiye O, van Gulik N, Allen J, Currey J. Importance of specific vital signs in nurses' recognition and response to deteriorating patients: a scoping review. *J Clin Nurs*. Jul 07, 2024;33(7):2544-2561. [doi: [10.1111/jocn.17099](https://doi.org/10.1111/jocn.17099)] [Medline: [38454551](#)]
17. Lee YJ, Cho KJ, Kwon O, Park H, Lee Y, Kwon J, 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*. Jun 2021;163:78-85. [[FREE Full text](#)] [doi: [10.1016/j.resuscitation.2021.04.013](https://doi.org/10.1016/j.resuscitation.2021.04.013)] [Medline: [33895236](#)]
18. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017. Presented at: NIPS '17; December 4-9, 2017:6000-6010; Long Beach, CA. URL: <https://dl.acm.org/doi/10.5555/3295222.3295349> [doi: [10.65215/ysbyhc05](https://doi.org/10.65215/ysbyhc05)]
19. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv*. Preprint posted online June 3, 2014. 2014. [[FREE Full text](#)] [doi: [10.3115/v1/d14-1179](https://doi.org/10.3115/v1/d14-1179)]
20. Aslam S, Rasool A, Li X, Wu H. CEL: a continual learning model for disease outbreak prediction by leveraging domain adaptation via elastic weight consolidation. *Interdiscip Sci*. Jun 28, 2025;17(2):390-408. [doi: [10.1007/s12539-024-00675-2](https://doi.org/10.1007/s12539-024-00675-2)] [Medline: [40019658](#)]
21. Rasool A, Tao R, Kashif K, Khan W, Agbedanu P, Choudhry N. Statistic solution for machine learning to analyze heart disease data. In: Proceedings of the 2020 12th International Conference on Machine Learning and Computing. 2020. Presented at: ICMLC '20; February 15-17, 2020:134-139; Shenzhen, China. URL: <https://dl.acm.org/doi/10.1145/3383972.3384061> [doi: [10.1145/3383972.3384061](https://doi.org/10.1145/3383972.3384061)]
22. Layeghian Javan S, Sepehri MM, Layeghian Javan M, Khatibi T. An intelligent warning model for early prediction of cardiac arrest in sepsis patients. *Comput Methods Programs Biomed*. Sep 2019;178:47-58. [doi: [10.1016/j.cmpb.2019.06.010](https://doi.org/10.1016/j.cmpb.2019.06.010)] [Medline: [31416562](#)]
23. Pan P, Wang Y, Liu C, Tu Y, Cheng H, Yang Q, et al. Revisiting the potential value of vital signs in the real-time prediction of mortality risk in intensive care unit patients. *J Big Data*. Apr 18, 2024;11(1):53. [doi: [10.1186/S40537-024-00896-8](https://doi.org/10.1186/S40537-024-00896-8)]
24. Areia C, King E, Ede J, Young L, Tarassenko L, Watkinson P, et al. Experiences of current vital signs monitoring practices and views of wearable monitoring: a qualitative study in patients and nurses. *J Adv Nurs*. Mar 15, 2022;78(3):810-822. [[FREE Full text](#)] [doi: [10.1111/jan.15055](https://doi.org/10.1111/jan.15055)] [Medline: [34655093](#)]
25. Eddahchouri Y, Peelen RV, Koeneman M, Touw HR, van Goor H, Bredie SJ. Effect of continuous wireless vital sign monitoring on unplanned ICU admissions and rapid response team calls: a before-and-after study. *Br J Anaesth*. May 2022;128(5):857-863. [[FREE Full text](#)] [doi: [10.1016/j.bja.2022.01.036](https://doi.org/10.1016/j.bja.2022.01.036)] [Medline: [35282866](#)]
26. Nyarko BA, Yin Z, Chai X, Yue L. Nurses' alarm fatigue, influencing factors, and its relationship with burnout in the critical care units: a cross-sectional study. *Aust Crit Care*. Mar 2024;37(2):273-280. [[FREE Full text](#)] [doi: [10.1016/j.aucc.2023.06.010](https://doi.org/10.1016/j.aucc.2023.06.010)] [Medline: [37580238](#)]
27. Teng Q, Liu Z, Song Y, Han K, Lu Y. A survey on the interpretability of deep learning in medical diagnosis. *Multimed Syst*. 2022;28(6):2335-2355. [[FREE Full text](#)] [doi: [10.1007/s00530-022-00960-4](https://doi.org/10.1007/s00530-022-00960-4)] [Medline: [35789785](#)]

28. Meng C, Trinh L, Xu N, Enouen J, Liu Y. Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. *Sci Rep.* May 03, 2022;12(1):7166. [[FREE Full text](#)] [doi: [10.1038/s41598-022-11012-2](https://doi.org/10.1038/s41598-022-11012-2)] [Medline: [35504931](#)]
29. Li X, Xiong H, Li X, Wu X, Zhang X, Liu J, et al. Interpretable deep learning: interpretation, interpretability, trustworthiness, and beyond. *Knowl Inf Syst.* Sep 14, 2022;64(12):3197-3234. [doi: [10.1007/S10115-022-01756-8](https://doi.org/10.1007/S10115-022-01756-8)]
30. Vinuesa R, Sirmacek B. Interpretable deep-learning models to help achieve the sustainable development goals. *Nat Mach Intell.* Nov 01, 2021;3(11):926. [doi: [10.1038/s42256-021-00414-y](https://doi.org/10.1038/s42256-021-00414-y)]
31. Zhang Y, Tino P, Leonardis A, Tang K. A survey on neural network interpretability. *IEEE Trans Emerg Top Comput Intell.* Oct 2021;5(5):726-742. [doi: [10.1109/tetci.2021.3100641](https://doi.org/10.1109/tetci.2021.3100641)]
32. Cuadrado D, Riaño D, Gómez J, Rodríguez A, Bodí M. Methods and measures to quantify ICU patient heterogeneity. *J Biomed Inform.* May 2021;117:103768. [[FREE Full text](#)] [doi: [10.1016/j.jbi.2021.103768](https://doi.org/10.1016/j.jbi.2021.103768)] [Medline: [33839305](#)]

## Abbreviations

**AUPRC:** area under the precision-recall curve

**AUROC:** area under the receiver operating characteristic curve

**CA:** cardiac arrest

**GRU:** gated recurrent unit

**HIPAA:** Health Insurance Portability and Accountability Act

**HR:** heart rate

**ICU:** intensive care unit

**LGBM:** light gradient boosting method

**SpO2:** oxygen saturation

**XGBoost:** extreme gradient boosting

Edited by A Schwartz, M Balcarras; submitted 03.Jun.2025; peer-reviewed by ♂ ♀, A Rasool; comments to author 16.Oct.2025; accepted 16.Dec.2025; published 09.Jan.2026

*Please cite as:*

Li Y, Lv L, Wang X

*Early Prediction of Cardiac Arrest Based on Time-Series Vital Signs Using Deep Learning: Retrospective Study*

*JMIR Form Res* 2026;10:e78484

URL: <https://formative.jmir.org/2026/1/e78484>

doi: [10.2196/78484](https://doi.org/10.2196/78484)

PMID:

©Yong Li, Lei Lv, Xia Wang. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 09.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.