Submitted: 22 August, 2023 Accepted: 28 November, 2023 Published: 08 April, 2024 DOI:10.22514/sv.2024.045

Open Access

Signa Vitae

ORIGINAL RESEARCH

The development and validation of a novel deep-learning algorithm to predict in-hospital cardiac arrest in ED-ICU (emergency department-based intensive care units): a single center retrospective cohort study

Yunseob Shin[1], Kyung-jae Cho[1], Mineok Chang[1], Hyun Youk[2], Yoon Ji Kim[3], Ji Yeong Park[3], Dongjoon Yoo[1,4,*]

[1]VUNO inc., 06541 Seoul, Republic of Korea
[2]Regional Trauma Center, Wonju Severance Christian Hospital, 26426 Wonju-si, Republic of Korea
[3]Yonsei University Wonju College of Medicine, 26426 Wonju-si, Republic of Korea
[4]Department of Critical Care Medicine and Emergency Medicine, Inha University Hospital, 22332 Incheon, Republic of Korea

***Correspondence**
dongjoon.yoo@vuno.co
(Dongjoon Yoo)

Abstract

Over recent years, the escalation of patient volumes in emergency departments (ED) worldwide has posed to the delivery of timely critical care. Intensive Care Unit (ICU) services became essential due to increasing acuity in EDs, and previous studies revealed a strong association between prolonged boarding times and unfavorable outcomes. Innovative strategies such as Emergency Department-based Intensive Care Units (ED-ICUs) have been introduced to optimize critical care delivery. Given the higher acuity and mortality rates in ED-ICU patients, the prediction of certain events, such as In-Hospital Cardiac Arrest (IHCA), has become abstruse. Conventional Early Warning Scores (EWSs) were developed to stratify the risk of conventional ICUs, but have never been validated in ED-ICU patients with higher acuity. Moreover, EWSs are predominantly focused on forecasting mortality and lack capability for real-time prediction. Our study aimed to develop and validate a deep-learning-based model to predict IHCA within 24 h in ED-ICU. We included 1975 patients admitted to ED-ICU. The study period was from 01 January 2019 to 31 December 2020. Our model, the Deep-ICU CMS (Central Monitoring System), uses four classic vital signs (blood pressure, heart rate, respiratory rate, and body temperature) as input. The model outperformed conventional EWSs in predicting IHCA and maintained performance even with extended prediction windows; it provided robust prediction within a 24-h window, setting it apart from models with restricted prediction horizons. It achieved notably high sensitivity and specificity, overcoming the alarm fatigue issue that is common in EWSs. This study pioneered IHCA risk stratification in ED-ICU and showcases Deep-ICU CMS as a robust prediction tool that overcomes the limitations of conventional EWSs. Prospective and external validation are now warranted to confirm the impact of Deep-ICU CMS in real-world practice. Given the scarcity of research in ED-ICU, our findings contribute valuable insights to optimizing critical care delivery.

Keywords

In-hospital cardiac arrest (IHCA); Emergency department-based intensive care unit (ED-ICU); Early warning score (EWS); Cardiac arrest (CA) prediction; Clinical deterioration; Machine learning; Deep learning; DeepCars

1. Introduction

Over recent years, emergency departments (ED) worldwide have witnessed a surge in patient volumes, leading to increased overcrowding and mounting challenges in the delivery of timely critical care [1]. Increasing acuity has led to a greater need for critical care services in EDs and intensive care units (ICUs) [2] and several studies have demonstrated that increased boarding time and ED crowding, which may lead to shortage of ED staff and the availability of resources to attend high acuity patients, is strongly associated with worse outcomes in critically ill patients [3–6]. The discharge of patients and movement to inpatients units represents an important output component of a conceptual framework to measure crowding within EDs that features three "buckets": input, throughput and output [7]. While the output represents a commonly cited reason for crowding in EDs [7, 8] during the COVID-19 pandemic, this challenge has been highlighted due to limited access block and the transmissibility of the coronavirus, thus leading to a dose-response increase in the rate of mortality in patients in the ED [9–11]. Novel strategies have been implemented to overcome this situation such as innovations in telemedicine, airway management using video laryngoscopes or infection control [12, 13]. Meanwhile, Emergency Department-based Intensive Care Units (ED-ICU) has been proposed to optimize the delivery of critical care and alleviate the crowding burden in EDs in contrast to conventional ICU admission systems [5, 14].

The South Korean government launched a similar program in 2004, designating 12 regional emergency medical centers across the country that featured 12 ED-ICU beds and was officially referred to as Emergency Intensive Care Units (EICU); these beds were dedicated solely to patients admitted from Eds [15]. In July 2023, 44 regional emergency medical centers with 20 ED-ICU beds were established throughout the country [16]. Although more than a decade has passed since the first implementation of ED-ICUs in Korea, only a few studies have been published on the nature of the population and the effects of implementation. Some studies found that ED-ICU patients had a higher severity at the time of admission and a higher mortality than conventional ICU-admitted patients [15, 17], making it more difficult to predict clinical deterioration, including In-Hospital Cardiac Arrest (IHCA), a crucial event that needs to be predicted and prevented; this condition is associated with an 18.8% rate of survival to hospital discharge after the event [18].

The information and cognitive load might cause safety issues in ICUs [19]. To ameliorate patient outcomes, EWSs were globally introduced as a uniform system to represent patient acuity. Widely used conventional EWSs in ICU settings, such as the Acute Physiology and Chronic Health Evaluation II (APACHE II) and Simplified Acute Physiology Score II (SAPS II), are mainly focused on predicting mortality itself; however, few researchers have investigated how these systems can be used to predict and prevent IHCA. Recently, a multicenter study demonstrated that conventional EWSs exhibited poor calibration, even though their prediction performance was acceptable [20]. Moreover, the limitation of most conventional EWSs used for prediction is that they are usually static, especially for use within 24 h of admission, and therefore do not reflect the highly variable acuity of patients who have received several treatments in real-time [21, 22]. Most importantly, to our knowledge, no study nor tool has stratified the risk of IHCA in ED-ICU settings, where patients tend to have higher acuity when admitted and the mortality differs from that in conventional ICUs [15, 17]. Given the scarcity of research in ED-ICU settings, our findings contribute valuable insights to optimizing the delivery of critical care for patients admitted from the ED. A brief comparative analysis with other conventional EWSs in conventional ICUs is presented in Table 1.

DeepCARS^{TM}, recently designated as a breakthrough device by the Food and Drug Administration (FDA), measures the risk of CA within 24 h of real-time vital sign observation. This system showed potential in predicting IHCA with higher sensitivity and a lower false-alarm rate than conventional EWSs during its original development in patients on a general ward [23]. In this study, we aimed to develop and validate a new deep-learning-based model for real-time prediction using the DeepCARS^{TM} engine and algorithm to predict IHCA within 24 h and stratify the risk of IHCA in ED-ICU patients.

2. Materials and methods

Wonju Severance Christian Hospital (WSCH) is a tertiary academic hospital comprising a regional emergency center and a regional trauma center that allocates 20 ED-ICU beds solely dedicated to patients admitted through the ED.

2.1 Study population

We conducted a retrospective cohort study using data collected from WSCH, which involved patients admitted to the ED-ICU at WSCH in South Korea over a two-year period, starting from 01 January 2019, and ending on 31 December 2020. We used data from 2019 to train our machine learning model, whereas data from the subsequent year were used for evaluation.

We followed specific exclusion criteria when selecting our study cohort. We excluded patients under the age of 20 years and those with a prior history of Out-of-Hospital Cardiac Arrest (OHCA) and IHCA before their ED-ICU admission. To maintain the integrity and relevance of our dataset, we excluded records that were entirely devoid of input feature values.

2.2 Data collection and preprocessing

During each patient's stay in the ED-ICU, we collected a comprehensive set of four classic vital signs: blood pressure (including systolic blood pressure (SBP) and diastolic blood pressure (DBP)), heart rate (HR), respiratory rate (RR), and body temperature (BT). In addition, patient age and the time of measurement for each input were obtained from electronic medical records (EMRs). To ensure the reliability and accuracy of our analysis, we marked any values that deviated extensively from the typical range or were non-numeric entries as missing and removed them from the analysis. Then we used a method called imputation to fill these gaps by replacing the missing values with the most recently recorded valid values. This approach helped keep our data complete and robust for analysis.

We also acquired the exact time of IHCA from the EMRs. Subsequently, we classified the data into two main types based on the IHCA status of patients. Samples pertaining to patients who experienced IHCA during their ED-ICU stay, ranging from the onset of the IHCA to 24 h prior, were labeled as "events". Conversely, samples associated with patients who did not encounter IHCA until their discharge from the ED-ICU were classified as "non-events".

2.3 Model development and validation

We constructed a deep learning-based predictive model called Deep-ICU CMS (Central Monitoring System) incorporating a vital sign encoder that consisted of a three bidirectional long short-term memory (LSTM) encoder, that is widely used

TABLE 1. Comparative analysis of risk stratification systems (SPTTS, APACHE II, NEWS and Deep-ICU CMS).

Feature	SPTTS	APACHE II	NEWS	Deep-ICU CMS
Outcome	Alert based on single abnormal vital sign	ICU risk stratification (mortality)	Early detection of clinical deterioration	Real-time prediction of IHCA within 24 hours
Setting	General wards	ICUs	General wards, Emergency department	ED-ICU in this study, possibly extendable to ICUs
Input variables	Classic 4 vital signs (BP, HR, BT, RR), mental status (AVPU)	Multiple physiological parameters, age, chronic health status (14 variables in total)	Six physiological parameters and whether patient is having oxygen therapy or not (7 variables on total)	Classic 4 vital signs (BP, HR, BT, RR), age, and their input time (6 variables)
Characteristics	Low performance and high false alarm. Lack of comprehensive understanding due to the nature of evaluating the input separately.	Static, reflecting the status within 24 hours after admission. Chronic health status is hard to be gathered automatically.	Intended to use in general wards. High false alarm.	Deep-learning based model.

SPTTS: Single-Parameter-Track-Trigger-System; APACHE II: Acute Physiology and Chronic Health Evaluation II; NEWS: National Early Warning Score; ICU: Intensive Care Unit; BP: blood pressure, HR: heart rate; RR: respiratory rate, BT: body temperature; CMS: Central Monitoring System; ED-ICU: Emergency Department-based Intensive Care Unit; AVPU: Alert, Verbal, Painful and Unresponsive.

in various tasks in the medical field [24, 25], and a binary classifier equipped with a fully connected layer. The LSTM encoder processes the recently recorded sequence of 20 vital signs. To prevent overfitting on the development dataset, dropout layers and batch normalization techniques were applied in addition to the LSTM encoder. The architecture of the Deep-ICU CMS is essentially the same as that of the model developed in a previous study by Kwon *et al*. [23] in 2019, which is used to predict IHCA in patients on general wards. LSTM, the key layer in the DeepCARS^{TM} and Deep-ICU CMS architectures for encoding the sequences of vital signs, is a type of neural network that features loops, thus allowing it to manage sequential data such as electronic health records. This structure aimed to mimic medical staff when reviewing the past medical information of patients when assessing their current condition. Based on the pretrained knowledge in the existing prediction model, we fine-tuned the model using ED-ICU data. More detailed explanations of the model architecture are provided in our previous research, including the outperforming prediction ability on general wards in a prospective multi-center study setting [23, 26–30].

To address the challenge of class imbalance, we adopted a data augmentation strategy by duplicating the data labeled as events, thereby adjusting the ratio of non-event-to-event instances during training. However, we refrained from using this strategy in the validation phase to evaluate the model's performance based on the original ratio. We experimented with various combinations of hyperparameters, including data batch size, learning rate, and the number of hidden dimensions for each layer. The combination that demonstrated the best performance was selected as the final model. We employed the Adam optimization algorithm [31] with a learning rate of 0.0001. Furthermore, our model was trained with a batch size of 256, and the LSTM encoder's hidden dimension was also configured to 256.

The training set included samples from patients admitted from 01 January 2019 to 31 December 2019, and the test set included samples from patients admitted from 01 January 2020 to 31 December 2020. Essentially, our dataset, sourced directly from electronic health records over two successive years, provided an in-depth snapshot of real-world ED-ICU scenarios. This solid dataset helped to enhance the external validity of the Deep-ICU CMS model. In addition, we utilized an early stopping method that stopped training at the optimal point based on the model's performance on the validation set. The variables utilized for training were the same as those used to evaluate the test set, as described earlier.

As previously detailed in our IHCA event labeling, the trained model predicts the likelihood of an IHCA occurring within the subsequent 24 hours from the present moment. If the model determines a substantial likelihood of an impending IHCA event, it produces a value approaching 100; conversely, a value nearing 0 indicates a minimal probability of the event occurring. In a practical EWS system, Deep-ICU CMS is designed to alert medical staff with a list of patients whose scores surpass the alarm threshold set by the professionals in the ICU. Furthermore, the medical team has the flexibility to adjust this threshold to manage the frequency of alarms.

2.4 Main outcome

The primary outcome of interest was the ability to predict the risk of IHCA within 24 h in an ED-ICU. IHCA was defined as the "cessation of cardiac activity, confirmed by the absence of a detectable pulse, unresponsiveness, and apnea", from the "in-hospital Utstein style" consensus guidelines published by the American Heart Association (AHA) [32]. We meticulously extracted and analyzed data related to IHCA incidents from multiple sources. This encompassed time-stamped orders of cardiopulmonary resuscitation (CPR) prescribed to patients, as well as the documentation of electrocardiographic (ECG) records associated with IHCA during attendance in the ED-ICU. Lethal rhythm (Pulseless Electric Activity (PEA) and Asystole) and shockable rhythm (pulseless Ventricular Tachycardia (pVT) and Ventricular Fibrillation (V.Fib) were both included. Then, we compared the predictive performance of our model with that of other conventional EWSs (the National Early Warning Score (NEWS) and single-parameter-Track-Tigger-System (SPTTS)). We excluded Do-Not-Resuscitate IHCA cases.

2.5 Secondary outcomes and statistical analysis

The performance of the IHCA prediction model was evaluated by comparison metrics from the receiver operating characteristic (ROC) curve, area under the curve (AUROC), and area under the precision-recall curve (AUPRC). The AUROC, a commonly used measure, illustrates the discriminatory ability of a model by plotting sensitivity against the false positive rate. In contrast, AUPRC addresses the issue of imbalanced data by quantifying the precision-sensitivity relationship. To conduct a robust comparison, we compared the Deep-ICU CMS AUROC and AUPRC values with three baseline methods: the NEWS, Logistic Regression (LR), and Random Forest (RF) methods. NEWS is an early warning system that is widely used in clinical practice, whereas the LR and RF models are machine-learning algorithms that are frequently used for their predictive capabilities.

In addition, we carried out comparative analysis with the NEWS tool by evaluating the F-score, net reclassification index (NRI), positive predictive value (PPV), negative predictive value (NPV), mean alarm count per day per 20 beds (MACPD), and the number needed to examine (NNE) at equivalent specificity levels as NEWS. The NNE is calculated as the total number of alarms divided by that of true positive alarms. The MACPD, which is calculated as the total number of alarms divided by that of days under the study period, and alarm rate, were compared at matching sensitivity levels, thus indicating that predictive performance and alarm rate are essential criteria for validating the practicality of an early warning system. A brief comparative analysis with conventional EWSs (APACHE II, NEWS) in terms of outcomes, input variables and characteristics are presented in Table 1 to demonstrate the differences between our model and existing methods. We did not compare our model with the APACHE II tool due to the unfairness of comparison with a static value that only reflects the status of early period of admission.

The potential of early IHCA prediction in the ED-ICU to enable the timely implementation of targeted interventions is remarkable and could potentially prevent the occurrence of IHCA. Such early identification and intervention hold promise for improving patient outcomes and alleviating the associated mortality burden. Hence, we evaluated the degree to which our model outperformed NEWS in the early prediction of IHCA in the ED-ICU.

We also analyzed additional results and the performance of our prediction model from various aspects, as described below.

2.5.1 Subgroup performance analysis

To understand the predictive performance within specific patient subgroups, we stratified patients based on sex, age, and risk level upon ED-ICU admission, as quantified by the APACHE II score. This subgroup analysis allowed us to thoroughly evaluate the performance of the Deep-ICU CMS model and its ability to accurately predict IHCA across different patient profiles.

2.5.2 Feature importance analysis

A crucial aspect of interpreting the decision-making process for the Deep-ICU CMS is determining the significance of individual vital sign characteristics. We used SHapley Additive exPlanations (SHAP) values to calculate the importance of each feature and time step. By quantifying the impact of specific vital signs on the predictions made by our model, we aimed to enhance the interpretability and understanding of the factors driving IHCA predictions.

2.5.3 Calibration analysis

A crucial aspect of the utility of a predictive model is its ability to provide accurate and reliable probability estimates. The calibration performance of the predictive model was assessed by focusing on its ability to produce well-calibrated output probabilities.

2.5.4 Statistical analysis

We conducted a comparative analysis between the AUROC scores of our proposed model and those of other baseline methods. DeLong's test was used to ascertain the statistical significance of the observed differences. All thresholds for determining statistical significance were set at $p < 0.05$. R software version 3.6.3 (R foundation, Vienna, Austria) was used for the analysis. T-tests were performed to verify the statistical significance of the differences in vital signs between the development and validation datasets. The missing value rate according to the time of model prediction was assessed in both the development and validation datasets; this strategy aimed to verify that the input chosen in our model was the same as that frequently employed in the real word, and that the scarcity of the missing value would have little influence on the results if implemented in real world. The threshold for determining statistical significance was set at $p < 0.05$. All analyses were conducted using Python (version 3.8.13) and the SciPy library (version 1.7.3).

3. Results

3.1 Baseline characteristics

The baseline characteristics of patients employed in the development (2019) and validation (2020) of our model are outlined in Table 2. The development dataset consisted of 970 admitted patients, with 103 experiencing IHCAs within their period of admission. The validation dataset, collected in the subsequent year, contained 1025 admitted patients, with 95 IHCAs reported. The construction of the development dataset and validation dataset is described as a flowchart in Fig. 1.

TABLE 2. Baseline characteristics of the development and validation dataset.

Baseline Characteristics	Development (2019)	Validation (2020)	*p*-value
Study period	2019-01-01–12-31	2020-01-01–12-31	
Total admission patients	970	1025	
IHCA patients within admissions	103	95	
Non-event patients/event patients	8.4	9.8	
Total vital sign records	109,859	110,508	
Vital sign records before IHCA within 24 h	3242	3306	
Non-event records/event records	32.8	32.4	
Male/Female	1.51	1.67	
Age (yr)	63.94 ± 15.68	63.87 ± 15.59	0.926
Length of stay in ED-ICU (day)	2.95 (1.58–5.73)	2.70 (1.28–5.63)	0.855
IHCA time after ED-ICU admission (h)	36.25 (8.05–99.17)	36.15 (12.13–107.08)	0.086
Total vital sign			
Systolic blood pressure (mmHg)	126.34 ± 24.58	127.98 ± 24.22	<0.001
Diastolic blood pressure (mmHg)	63.67 ± 13.34	64.96 ± 14.45	<0.001
Heart rate (/min)	89.13 ± 21.55	86.52 ± 20.21	<0.001
Respiratory rate (/min)	20.62 ± 6.00	18.68 ± 5.82	<0.001
Body temperature (ºC)	36.98 ± 0.55	36.90 ± 0.56	<0.001
Vital sign within 24 hours before IHCA			<0.001
Systolic blood pressure (mmHg)	101.66 ± 29.86	100.93 ± 28.61	<0.001
Diastolic blood pressure (mmHg)	54.31 ± 13.53	54.42 ± 13.54	<0.001
Heart rate (/min)	107.04 ± 27.17	102.68 ± 25.85	<0.001
Respiratory rate (/min)	24.29 ± 6.29	23.57 ± 7.10	<0.001
Body temperature (ºC)	36.95 ± 0.80	36.56 ± 0.74	<0.001
Measurement interval of vital signs			
Systolic blood pressure (h)	0.89	0.99	
Diastolic blood pressure (h)	0.89	0.99	
Heart rate (h)	0.90	0.99	
Respiratory rate (h)	0.93	1.04	
Body temperature (h)	1.68	1.71	

IHCA: In-Hospital Cardiac Arrest; ED-ICU: Emergency Department-based Intensive Care Unit.

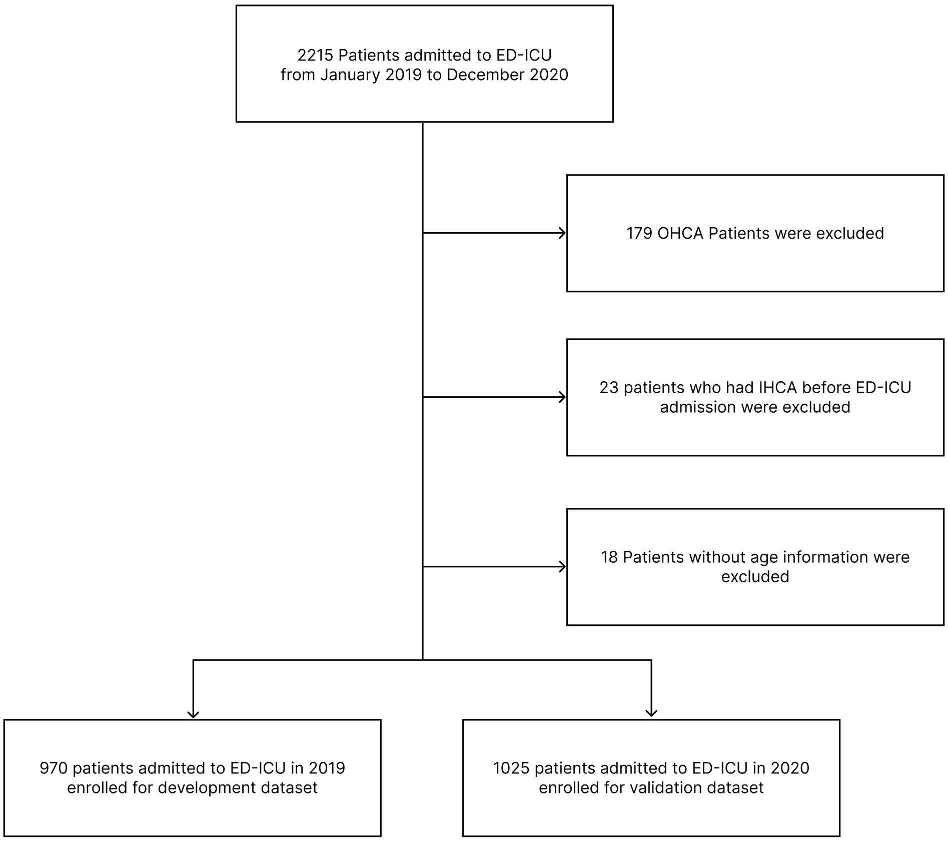

FIGURE 1. A flowchart showing the exclusion and inclusion criteria applied during patient selection. ED-ICU: Emergency Department-based Intensive Care Unit; OHCA: Out-of-Hospital Cardiac Arrest; IHCA: In-Hospital Cardiac Arrest.

In terms of demographic data, the ratio of male to female patients was slightly higher in the validation dataset (1.67) than in the development dataset (1.51). The mean age was similar in both datasets (approximately 64 years), with a slight standard deviation around 15.6 years. The mean length of stay in the ED-ICU was marginally shorter in the validation dataset (2.70 days) than in the development dataset (2.95 days). The median time to IHCA after admission to the EDICU was almost identical between the two datasets.

There were slight differences between the development and validation datasets for certain baseline patient characteristics. However, key parameters, such as the ratio of male to female patients, age, and length of stay in the ED-ICU, remained consistent. Interestingly, both datasets demonstrated marked shifts in vital signs within the 24 h preceding an IHCA event. Table 3 illustrates the missing rate of vital signs in the validation dataset over different time intervals. As the interval increased, the missing data rate generally decreased for all vital signs. Notably, BT had a significantly higher missing rate at the 1-hour interval (36.39%) when compared to other vital signs, although this decreased substantially over longer intervals. A similar trend was observed in the development dataset, as shown in Table 4.

3.2 Predictive performance

As illustrated in Fig. 2 and Table 5, our predictive model showed exceptional capabilities for the prediction of IHCA within 24 h in the ED-ICU environment, markedly outperforming the three baseline methods (NEWS, LR and RF). Our model demonstrated a robust predictive performance, with an AUROC score of 0.923 (95% confidence interval (CI), 0.919–0.929). This considerably overshadowed LR, with an AUROC of 0.882 (95% CI, 0.879–0.887); RF, with an AUROC of 0.881 (95% CI, 0.876–0.887); and NEWS, with an AUROC of 0.864 (95% CI, 0.860–0.871). Similarly, our model surpassed the three baselines when evaluated by AUPRC. Our model had an AUPRC of 0.4068 (95% CI, 0.3970–0.4272), LR had an AUPRC of 0.2925 (95% CI, 0.2824–0.3084), RF had an AUPRC of 0.2778 (95% CI, 0.2684–0.2970), and NEWS had an AUPRC of 0.2908 (95% CI, 0.2800–0.3039). The higher AUROC and AUPRC values of the Deep-ICU CMS indicated that in addition to accurate prediction, the model is particularly skilled at distinguishing between patients who will have an IHCA and those who will not, even when positive cases are scarce. These findings highlight the superior predictive strength of our model for predicting the risk of IHCA within 24 h, signifying its potential applicability in a clinical context.

3.3 Performance according to different event times

We evaluated the performance of our model, along with those of LR, RF and NEWS, at multiple timeframes preceding IHCA events (3, 6, 12 and 24 h). Table 6 shows that our model had the highest AUROC value of 0.9473, exceeding those of LR (0.9175), RF (0.9249) and NEWS (0.9166) 3 h ahead of IHCA. Six hours before the event, our model continued to exhibit

TABLE 3. Missing rate of data in the validation dataset with regards to time interval.

Interval (h)	SBP missing rate (%)	DBP missing rate (%)	HR missing rate (%)	RR missing rate (%)	BT missing rate (%)
1	0.79	0.77	0.83	1.83	36.39
2	0.71	0.71	0.75	1.48	4.57
3	0.54	0.55	0.60	1.31	1.37
6	0.25	0.24	0.40	0.84	0.64
24	0.04	0.04	0.16	0.41	0.28
For all input data	1.81	1.72	2.86	7.58	41.87

SBP: systolic blood pressure; DBP: diastolic blood pressure; HR: heart rate; RR: respiratory rate; BT: body temperature.

TABLE 4. Missing rate of data in the development dataset with regards to time interval.

Interval (h)	SBP missing rate (%)	DBP missing rate (%)	HR missing rate (%)	RR missing rate (%)	BT missing rate (%)
1	0.17	0.18	0.93	1.25	41.46
2	0.08	0.08	0.95	1.15	5.89
3	0.05	0.04	0.93	1.15	1.58
6	0.04	0.04	0.90	1.18	1.33
24	0.04	0.04	0.78	1.00	0.87
For all input data	1.19	1.16	2.71	6.08	46.12

SBP: systolic blood pressure; DBP: diastolic blood pressure; HR: heart rate; RR: respiratory rate; BT: body temperature.

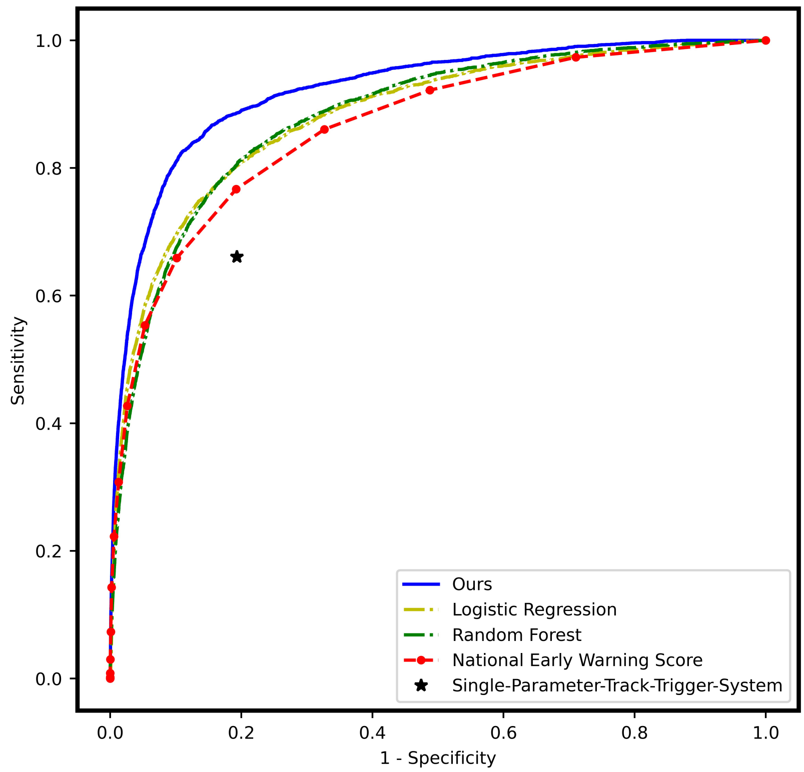

FIGURE 2. Area Under the Receiver Operating Characteristics curve. Our model outperformed Logistic Regression (LR), Random Forest (RF), NEWS (National Early Warning Score), and SPTTS (Single-Parameter-Track-Trigger-System) in all ranges of false positive rate (FPR).

TABLE 5. Overall performance comparison of IHCA prediction within 24 h measured by AUROC and AUPRC scores.

Prediction model	Total AUROC (95% CI)	Total AUPRC (95% CI)
Deep-ICU CMS	0.923 (0.919–0.929)	0.4068 (0.3970–0.4272)
Logistic Regression	0.882 (0.879–0.887)	0.2925 (0.2824–0.3084)
Random Forest	0.881 (0.876–0.887)	0.2778 (0.2684–0.2970)
National Early Warning Score (NEWS)	0.864 (0.860–0.871)	0.2908 (0.2800–0.3039)
Single-Parameter-Track-Trigger-System	0.734 (0.729–0.744)	0.0732 (0.0715–0.0787)

AUROC: Area Under Receiver Operating Characteristics; AUPRC: Area Under Precision-Recall Curve; CI: confidence interval; ICU: Intensive care unit; CMS: Central Monitoring System.

TABLE 6. Comparison of AUROC scores over time before cardiac arrest for different prediction models.

Time before IHCA (h)	Deep-ICU CMS	LR	RF	NEWS	SPTTS
2	0.949	0.920	0.926	0.919	0.811
3	0.947	0.918	0.925	0.917	0.812
6	0.941	0.912	0.915	0.906	0.791
12	0.930	0.898	0.899	0.881	0.753
24	0.923	0.882	0.881	0.864	0.734

CA: Cardiac Arrest; LR: Logistic Regression; RF: Random Forest; NEWS: National Early Warning Score; SPTTS: Single-Parameter-Track-Trigger-System; IHCA: In-Hospital Cardiac Arrest; ICU: Intensive care unit; CMS: Central Monitoring System.

superior performance, registering an AUROC of 0.9405 as opposed to LR (0.9116), RF (0.9150) and NEWS (0.9059). Furthermore, out model achieved the highest AUROC even at broader intervals prior to IHCA. Twelve hours before the event, our model's AUROC was 0.9297, thus outperforming LR (0.8980), RF (0.8989), and NEWS (0.8814). In addition, twenty-four hours before IHCA, our model outperformed the others, with an AUROC of 0.9225; in comparison, the AUROC values for LR, RF and NEWS were 0.8822, 0.8814 and 0.8645, respectively. Overall, our model consistently exhibited superior predictive performance over LR, RF and NEWS, irrespective of the timeframe leading up to the IHCA events. This highlights the fact that Deep-ICU CMS consistently outperformed other models in predicting IHCAs across all examined timeframes, signifying its robustness and potential for early clinical intervention. Notably, its marked advantage over single-parameter systems such as Single-Parameter-Track-Trigger-System (SPTTS) emphasizes the value of comprehensive data-driven models in critical care settings.

3.4 Comparative analysis with NEWS and alarm performance

Next, we performed in-depth comparative analysis with NEWS and evaluated key metrics at the same specificity levels and analyzed alarm performance across multiple cutoff points at the same sensitivity levels. The comprehensive outcomes of these assessments are presented in Tables 7 and 8. Classification outcomes, including those from the confusion matrix, were evaluated each time a new vital sign was recorded. In IHCA patients, measures were taken specifically within the 24 hours before the IHCA event. In contrast, in normal patients, we considered vital signs throughout the entire period of admission. Finally, the overall sensitivity and specificity were calculated using the accumulated true positives, true negatives, false positives, and false negatives from the entire validation set.

When comparing our model to NEWS, we found that our model consistently improved the sensitivity, NRI, and both predictive values (PPV and NPV) while maintaining similar alarm counts (MACPD) and achieving lower NNE values. For instance, at the NEWS sensitivity levels of 0.973 and 0.922, our model improved the sensitivity to 0.990 and 0.964, respectively. Simultaneously, our model increased NRI values, demonstrating a superior ability to augment IHCA predictions. These improvements were also mirrored in the predictive values, reinforcing the precision of our model for predicting IHCA and non-arrest events.

In terms of alarm performance, our model demonstrated superior effectiveness when compared with NEWS across all of the sensitivity levels examined. For instance, at sensitivity levels of 0.973 and 0.922, our model yielded a higher specificity, a reduced NNE, and a lower MACPD, translating to substantial MACPD reduction rates of 22.5% and 40.2%, respectively. This pattern of improved performance was extended to lower sensitivity levels, with our model persistently achieving higher specificity, reduced NNE and lower MACPD, thus highlighting its practicality and efficiency in clinical settings.

Thus, our model demonstrated enhanced performance compared with NEWS, both in terms of classification accuracy and alarm efficiency, making it a promising alternative for the early detection of IHCA in clinical settings.

TABLE 7. Comparative alarm efficiency of NEWS and our model at equivalent sensitivity levels.

Cutoff	Sensitivity	Specificity	NNE	MACPD	Reduction rate of MACPD
NEWS $\geq$1	0.973	0.290	24.658	311	
Deep-ICU CMS $\geq$8.9	0.973	0.436	19.798	241	−22.5%
NEWS $\geq$2	0.922	0.512	18.155	214	
Deep-ICU CMS $\geq$23.9	0.922	0.717	10.961	128	−40.2%
NEWS $\geq$3	0.861	0.673	13.318	148	
Deep-ICU CMS $\geq$45.5	0.861	0.850	6.638	77	−48.0%
NEWS $\geq$4	0.767	0.808	9.132	95	
Deep-ICU CMS $\geq$68.9	0.767	0.921	4.351	50	−47.4%
NEWS $\geq$5	0.659	0.898	6.000	58	
Deep-ICU CMS $\geq$85.0	0.659	0.954	3.278	35	−39.7%
NEWS $\geq$6	0.554	0.947	4.107	37	
Deep-ICU CMS $\geq$92.9	0.554	0.972	2.669	26	−29.7%
NEWS $\geq$7	0.427	0.974	2.989	24	
Deep-ICU CMS $\geq$97.1	0.427	0.985	2.159	18	−25.0%
NEWS $\geq$8	0.308	0.987	2.366	15	
Deep-ICU CMS $\geq$98.7	0.308	0.993	1.743	12	−20.0%

NEWS: National Early Warning Score; MACPD: Mean Alarm Count per Day; NNE: Number needed to examine; ICU: Intensive care unit; CMS: Central Monitoring System.

TABLE 8. Comparative metrics of NEWS and our model at equivalent specificity levels.

Cutoff	Sensitivity	Specificity	PPV	NPV	*F*-score	MACPD	NNE	NRI
NEWS $\geq$1	0.973	0.291	0.041	0.997	0.078	311	24.658	-
Deep-ICU CMS $\geq$5.6	0.990	0.291	0.041	0.999	0.079	300	24.190	0.006
NEWS $\geq$2	0.922	0.512	0.055	0.995	0.104	214	18.155	-
Deep-ICU CMS $\geq$11.4	0.964	0.512	0.057	0.998	0.108	210	17.410	0.396
NEWS $\geq$3	0.861	0.674	0.075	0.994	0.138	148	13.318	-
Deep-ICU CMS $\geq$20.1	0.932	0.674	0.081	0.997	0.149	145	12.321	0.568
NEWS $\geq$4	0.767	0.808	0.110	0.991	0.192	95	9.132	-
Deep-ICU CMS $\geq$36.7	0.885	0.808	0.125	0.996	0.218	93	8.031	0.693
NEWS $\geq$5	0.659	0.898	0.167	0.988	0.266	58	6.000	-
Deep-ICU CMS $\geq$59.9	0.812	0.898	0.197	0.994	0.317	59	5.079	0.761
NEWS $\geq$6	0.554	0.947	0.243	0.986	0.338	37	4.107	-
Deep-ICU CMS $\geq$81.7	0.681	0.947	0.284	0.990	0.401	38	3.521	0.781
NEWS $\geq$7	0.427	0.974	0.335	0.982	0.375	24	2.989	-
Deep-ICU CMS $\geq$93.9	0.533	0.974	0.388	0.985	0.449	24	2.578	0.778
NEWS $\geq$8	0.308	0.987	0.423	0.979	0.356	15	2.366	-
Deep-ICU CMS $\geq$97.6	0.401	0.987	0.489	0.982	0.441	16	2.044	0.765

NEWS: National Early Warning Score; PPV: Positive Predicted Value; NPV: Negative Predicted Value; MACPD: Mean Alarm Count per Day; NNE: Number needed to examine; NRI: Net reclassification index; ICU: Intensive care unit; CMS: Central Monitoring System.

TABLE 9. Comparative time-to-event analysis between our model and NEWS.

Cumulative percentage of predicted CA patients	Deep-ICU CMS (h)	NEWS (h)	Difference of predicted time (h)
26%	23.2	20	3.2
35%	20.0	16	4.0
44%	15.7	12	3.7
55%	10.8	8	2.8
65%	6.4	4	2.4
71%	5.2	0	5.2

CA: Cardiac Arrest; NEWS: National Early Warning Score; ICU: Intensive care unit; CMS: Central Monitoring System.

3.5 Early prediction superiority

Our model consistently demonstrated a superiority over the NEWS in terms of the early prediction of IHCA within the ED-ICU. We set the alarm threshold of our model to the same specificity as when the NEWS was greater than or equal to 8. As shown in Table 9, our model provided warnings 23.2 h before the event for the first 26% of the predicted IHCA cases; this was 3.2 h ahead of the 20 h warning provided by NEWS. This advantage increased with the cumulative percentage of patients with predicted IHCA. In the first 35% of cases, our model delivered warnings 20 h in advance, a full 4 h earlier than NEWS. Similarly, in the first 44% and 55% of predicted cases, our model achieved a lead time of 3.7 and 2.8 h earlier than NEWS, providing alerts 15.7 and 10.8 h before the actual IHCA event, respectively. The disparity between our model and NEWS was most pronounced in the top segments of the predicted IHCA cases. For the first 65% and 71% of the predicted cases, our model signaled warnings 6.4 and 5.2 h prior to IHCA, thus outperforming NEWS by a substantial 2.4 and 5.2 h, respectively. We also found that our model predicted IHCA 14.4 h earlier on average.

In summary, our model demonstrated remarkable superiority over NEWS for the early prediction of IHCA. By offering alerts earlier across a broad spectrum of patient populations, our model demonstrated the potential to considerably enhance the window for effective and timely interventions in ED-ICUs.

3.6 Subgroup performance analysis

We conducted subgroup analysis to evaluate the performance of our model across various demographic and clinical groups by comparing it with LR, RF and NEWS.

When examining patients grouped by age, as shown in Table 10, our model demonstrated a higher area under the curve (AUC) for each age category than LR, RF and NEWS. Among patients aged 40–60 years, the AUC of our model was 0.960, thus surpassing that of the other models. Similarly, superior performance was observed in patients aged 60–80 years and in those aged 80 years or above, with AUC values of 0.912 and 0.910, respectively. We also analyzed performance based on sex. As shown in Table 9, our model exhibited higher AUC values for both the male and female subgroups, at 0.919 and 0.929, respectively, thus outperforming LR, RF and NEWS in both cases. Finally, we assessed the results using the APACHE II score, a widely used classification for the severity of disease. As shown in Table 9, our model exhibited superior AUC values for all categories of APACHE II scores. In this experiment, we only utilized the prediction results from patients with a valid APACHE II score. For scores between 0 and 15, 15 and 25, and above 25, our model achieved AUC values of 0.934, 0.900 and 0.932, respectively, thus surpassing those of the other models for each score category.

In summary, our model consistently outperformed LR, RF and NEWS across various subgroups, demonstrating its robustness and potential for broader applications in different patient populations.

3.7 Feature importance analysis

In our feature importance analysis, we utilized the Shapley Additive exPlanations (SHAP) framework, which provides a unified measure of feature importance that appropriately allocates the contribution of each feature to the model output.

As shown in Fig. 3A, the vital signs that emerged as the most critical in the performance of our model, ranked in descending order of importance, were SBP, HR, RR, BT and DBP. This highlights the significance of these physiological markers for the early prediction of IHCA. In addition, our temporal analysis highlighted the importance of recent data for driving model predictions.

Moreover, our analysis emphasized the importance of applying the most recent data points in the model's predictions. As shown in Fig. 3B, the SHAP values associated with the sequences increased as the observations approached the present, indicating that the most recent data exerted a stronger influence on the prediction outcome. This finding highlighted the dynamic nature of patient vital status and the necessity for current information to accurately predict IHCA.

3.8 Calibration analysis

Calibration is a critical aspect of the performance of a prediction model, as it measures the agreement between the predicted probabilities of an event and the observed frequencies. In this analysis, we used the expected calibration error (ECE) loss to quantify the calibration performance of our model and the NEWS. Our model demonstrated an extensively lower loss of ECE (0.029) than NEWS (0.248), indicating that our model predictions were more aligned with the observed frequencies of IHCA. A lower loss of ECE signified a smaller discrepancy between the model's predicted probabilities and actual outcomes, thus indicating better calibration.

The superior calibration performance of our model is il-

TABLE 10. Subgroup performance analysis.

Group by age	Number of CA patients	Number of normal patients	Deep-ICU CMS	LR	RF	NEWS
$20 \leq$ age < 40	0	47	-	-	-	-
$40 \leq$ age < 60	20	208	0.960	0.906	0.908	0.894
$60 \leq$ age < 80	44	418	0.912	0.888	0.875	0.867
$80 \leq$ age	31	257	0.910	0.855	0.860	0.832
Group by gender	Number of CA patients	Number of normal patients	Deep-ICU CMS	LR	RF	NEWS
Male	579	62	0.919	0.891	0.886	0.872
Female	351	33	0.929	0.863	0.871	0.851
Group by APACHE II score	Number of CA patients	Number of normal patients	Deep-ICU CMS	LR	RF	NEWS
$0 \leq$ score < 15	21	407	0.934	0.889	0.905	0.863
$15 \leq$ score < 25	39	363	0.900	0.853	0.853	0.829
$25 \leq$ score	23	75	0.932	0.897	0.867	0.896
$15 \leq$ score < 25	39	363	0.900	0.853	0.853	0.829
$25 \leq$ score	23	75	0.932	0.897	0.867	0.896

LR: Logistic Regression; RF: Random Forest; NEWS: National Early Warning Score; CA: Cardiac Arrest; ICU: Intensive care unit; CMS: Central Monitoring System.

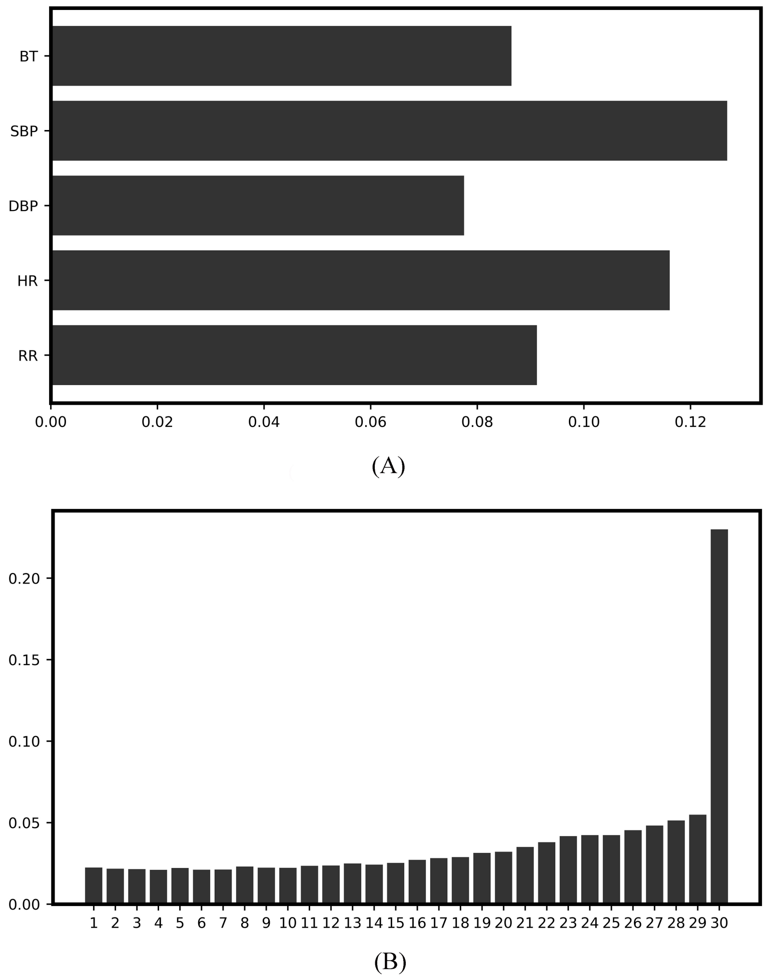

FIGURE 3. Absolute SHapley Additive exPlanations (SHAP) values of (A) each vital sign and (B) each sequence. BT: body temperature; SBP: systolic blood pressure; DBP: diastolic blood pressure; HR: heart rate; RR: respiratory rate.

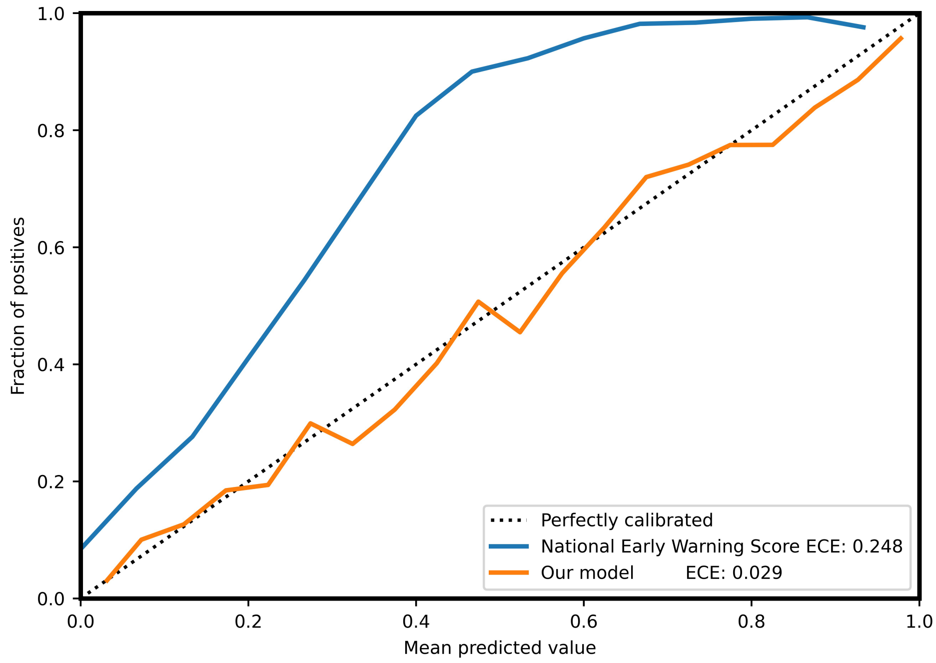

FIGURE 4. Comparative calibration analysis between the National Early Warning Score (NEWS) and our model. ECE: expected calibration error.

lustrated in Fig. 4, which visually highlights its improved alignment with the observed outcomes. This evidence signifies that our model provides more reliable and trustworthy probability estimates for IHCA in ED-ICU settings than NEWS. Therefore, improved calibration can enhance clinical decision making by offering accurate risk estimation, thus enabling timely interventions to prevent IHCA.

4. Discussion

In this study, we developed and validated a novel real-time prediction score (Deep-ICU CMS) to stratify the risk of IHCA in 24 h for patients admitted to ED-ICUs. The Deep-ICU CMS model had a higher discriminating performance than the commonly used EWSs, NEWS and SPTTS, with an AUROC score of 0.923 (95% CI, 0.919–0.929) compared to an AUROC of 0.864 (95% CI, 0.860–0.871), along with an AUROC of 0.734 (95% CI, 0.729–0.744) for the prediction of IHCA within 24 h (Table 5 and Fig. 2). This high prediction performance was maintained even when the prediction time was shortened to 2 h (Table 6). Higher specificity was noted, with the same sensitivity as NEWS, for all cutoff scores; this means that lower alarms were produced regardless of any sensitivity cutoff. The reduction rate ranged from 20.0% to 48.0% when compared to NEWS, which overcame the inherent limitations of conventional EWSs which tend to have a high false alarm rate [33, 34] and can also lead to alarm fatigue in the medical staff in charge and ultimately a desensitization to false alarms (Table 7) [35]. Attenuating alarm fatigue may rescue medical staff who are responsible for critical care from information that can be overlooked in ICUs. Furthermore, higher sensitivity was demonstrated at the same specificity level (Table 7). Calibration performance, along with discrimination performance, which should be considered when judging model accuracy [36, 37], significantly higher than that of NEWS (Fig. 4). The overall performance, including AUROC, specificity, false alarm rate, sensitivity, and calibration of the Deep-ICU CMS for predicting IHCA in 24 h for ED-ICU patients, outperformed that of NEWS.

To the best of our knowledge, this study is the first to stratify the risk of IHCA in patients in ED-ICUs, especially in real-time. Global studies relating to ED-ICUs are sparse as this is a relatively new system that has been implemented for only one or two decades. In South Korea, patients admitted to the ED-ICU tend to have higher acuity and mortality rates than those admitted to formal ICUs [17]. Early-goal directed therapy and reassessment have been emphasized over recent years to reduce mortality in critically ill patients [38–40]. The consensus on the need for timely reassessments through the continuous monitoring of critical patients is crucial if we are to improve survival, and safety; this strategy is becoming widely accepted [41], while an increased number of monitored parameters leads to complexity in terms of interpretation and the burden of documentation involved [42, 43]. Despite the accuracy of laboratory blood tests, which can influence up to 70% of diagnostic or treatment decisions, unnecessary redundant and repeated laboratory tests have recently been raised as a concern [44]. On the other hand, vital signs are recorded every hour at a minimum without any extra cost in most ICUs. Moreover, in terms of real-time prediction models, vital sign-based models tend to be more appropriate owing to the inherent nature of short-term interval input time, therefore providing prediction scores at least on an hourly basis.

The conventional EWSs that are widely used in clinical practice to predict patient acuity, such as the APACHE II and SAPS II, are usually static, use data within 24 h of admission, lack the ability to predict patient acuity in real- time, and do not reflect the change in a patient's condition through admission and treatment [45]. Regardless of their prediction performance, the poor calibration of conventional EWSs has been reported in several recent studies [36, 37, 46].

To our knowledge, no model has attempted the risk stratification of IHCA for ED-ICU patients, and few models have tried to predict IHCA in ICU patients based on machine learning. Most previous researchers chose to compromise the prediction window by shortening the prediction time to guarantee prediction accuracy, as prediction performance (Compared in Table 11) [47–49], otherwise referred to as prediction error, falls dramatically when the time of prediction, the so-called "prediction horizon" increases, even if a deep-learning technique is applied [50–52]. Our model maintains robust performance with a 24 h prediction window, meaning that it can predict targeted events during the next 24 h if clinically implemented; this differs from other models that only preserve prediction performance when predicting events during the next few hours.

Another means of improving prediction performance is to provide diverse information to the model, including laboratory tests, vast nursing records, and other data that are not always available in real clinical practice [47, 53, 54]. Most model-development protocols are based on retrospective studies and multivariate inputs, and are therefore unable to evade the problem of missing values. Several guidelines, including STrengthening the Reporting of OBservational studies in Epidemiology (STROBE), have been published to correct statistical errors in retrospective cohort studies; however, many of these guidelines have been overlooked [55, 56]. Missing data can lead to low performance but can also distort the entire study through selection bias, which can potentially invalidate the entire study [55, 57]. Our model used only the four classic vital signs, the age of patients and their time of measurements; these are parameters that are rarely missing in the ICU or general ward. Consequently, this strategy is expected to have the same outstanding performance when compared to other models developed to date when implemented in real-world practice [45, 47–49].

Another important consideration during model development was the selection of outcomes to predict clinical deterioration. Controversies remain with regards to defining clinical deterioration, and several studies have defined IHCA as the most important outcome, with no objections to its inclusion in the clinical deterioration criteria [58]. Previous large randomized studies across the world have chosen complex outcomes, including IHCA, unplanned ICU transfer (UIT) and mortality, to evaluate the effect of Rapid Response System (RRS) implementation due to the low number of patients anticipated to reach the individual components of this endpoint [59]. This trend continued for various reasons: for example, death and UIT cases are well clarified and defined in most datasets and contribute to a large portion of the prediction performance in both conventional and deep-learning-based EWSs [60, 61]. Our model was intended to focus only on IHCA cases while maintaining a high degree of robustness.

Nevertheless, a dramatic improvement in the mortality of critically ill patients has been demonstrated since its implementation. ICU mortality remains high, ranging from 13% to 20%, as reported in recent studies from the United States and Europe [62, 63]. Financially, the US spends approximately $82 billion annually on ICU admissions, accounting for approximately 0.66% of the US gross domestic product (GDP) [64]. With appropriate staff, monitoring and treatment, it is possible to save almost $13 million ICU costs on an annual basis [65]. Our model also alleviates the financial burden on patients by implementing an ED-ICU system.

If implemented in a real-world ED-ICU, Deep-ICU CMS will help medical staff to stratify IHCA risks in real-time. This

TABLE 11. Comparative analysis of prediction models in recent studies.

Prediction model	Year of publication	Target prediction	Data source	Input features	Prediction horizon	AUROC performance
Deep-ICU CMS	-	IHCA	ED-ICU of Wonju Severance Christian Hospital	Four vital signs	24 h	0.923
Sung *et al*. [47]	2021	Multiple events (mortality, sepsis, AKI)	ICU of the National Health Insurance Corporation Ilsan Hospital	Five vital signs, 10 laboratory results, GCS	12 h (mortality, AKI), 6 h (sepsis)	0.938 (mortality), 0.738 (sepsis), 0.760 (AKI)
Kim *et al*. [48]	2020	IHCA	ICU of the Asan Medical Center	Vital signs, laboratory results, SOFA score	24 h, 48 h	0.875 (24 h), 0.841 (48 h)
Yijing *et al*. [49]	2022	IHCA	MIMIC-III dataset	Four vital signs	2 h	0.94

AUROC: Area Under Receiver Operating Characteristics; IHCA: In-hospital Cardiac Arrest; AKI: Acute Kidney Injury; GCS: Glasgow Comma Scale; SOFA: Sequential Organ Failure Assessment; ICU: Intensive care unit; CMS: Central Monitoring System; ED-ICU: emergency department-based intensive care unit; MIMIC: Medical Information Mart for Intensive Care.

stratification will help medical staff to identify patients with the highest acuity and provide them with the optimal resuscitation method, including more invasive procedures such as Continuous Renal Replacement Therapy (CRRT) and Extracorporeal Membrane Oxygenation (ECMO). In addition, the model may provide insights that prompt review and adjustment of the initial resuscitation strategy. The improvement of mortality would also be anticipated by preventing IHCA. The prompt stabilization of critically-ill patients with Early Goal Directed Therapy (EGDT) is expected to reduce ICU stays and financial burden. The faster turn-over of ED-ICU beds will help to mitigate ED overcrowding leaving room to faster admission from ED to ICU. Based on the AHA, 48% of the surveyed hospitals in the US were deficient in terms of the number of critical care staff [66]. In a recent paper, one proposed way to solve this problem was the adoption of artificial intelligence (AI) to assist information processing and the facilitation of treatment [67]. Deep-ICU CMS is poised to meet these prerequisites, given its demonstrated proficiency in alarm reduction and IHCA prediction in contrast with conventional methods.

Our study and the model itself have some limitations, especially with regards to implementation in the real-word. First, the study was conducted retrospectively in a tertiary single center; hence, the results need to be validated externally in other ED-ICU centers, to reduce bias. A well-designed multicenter and retrospective study should be conducted to further investigate patient safety and efficacy when implemented in the real-world. Second, it is necessary to conduct a meticulously planned prospective clinical trial to provide additional evidence for the efficacy of the Deep-ICU CMS^{TM} as a screening tool in clinical practice. This prospective trial should seek to demonstrate overall ED-ICU mortality improvement following implementation of the model. Third, our findings were derived from a solitary tertiary care hospital that included a regional emergency center with a high level of acuity, but included a relatively small number of patients. Consequently, it may be unreasonable to anticipate comparable advantages when implementing the Deep-ICU CMS^{TM} in all hospitals. Therefore, the generalizability of our results is limited.

5. Conclusions

The ED-ICU is a relatively new approach aimed at improving patient outcomes. Nonetheless previous studies have shown advancements in the reduction of mortality rates; furthermore, there is notable potential for enhancing patient outcomes and ensuring patient safety. There is a lack of risk stratification methods that are relevant for patients admitted to ED-ICUs. This is the first study to report a model for predicting IHCA in an ED-ICU. A higher predictive power was evident when compared to conventional EWSs and artificial intelligence-based models for the prediction of ICU-admitted patients, even with longer prediction windows and low input features. Given the scarcity of research in ED-ICU settings, our findings contribute valuable insights to the optimization of critical care delivery for patients admitted from EDs.

The optimization of initial resuscitation, dealing with the overcrowding and shortage of EDs and ICUs, including medical staff, has become increasingly problematic of late. Deep-ICU CMS is poised to meet the prerequisites and yield better outcomes and safety measures for critically ill patients admitted from EDs with ED-ICU systems.

Our research has several limitations, due to it being a retrospective study based in a single center. Before being implemented in the real world, a well-designed multicenter retrospective study for external validation should be undertaken to seek external validation and reduce bias. Furthermore, a well-designated multicenter prospective study, including external validation, is expected to provide additional evidence and clinical impact in real-world practice.

ABBREVIATIONS

AHA, American Heart Association; APACHE II, Acute Physiology and Chronic Health Evaluation II; AUPRC, Area Under Precision-Recall Curve; AUROC, area under the receiver operating characteristics; BT, body temperature; CPR, cardiopulmonary resuscitation; DBP, diastolic blood pressure; ECE, expected calibration error; ECG, electrocardiographic; ED-ICU, emergency department-based intensive care unit; EGDT, Early Goal Directed Therapy; EMRs, electric medical records; EWS, early warning score; FDA, Food and Drug Administration; HR, heart rate; IHCA, in-hospital cardiac arrest; LR, logistic regression; LSTM, long short-term memory; MACPD, mean alarm count per day; NEWS, national early warning score; NNE, number needed to examine; NRI, net reclassification index; NPV, negative predictive value; PEA, pulseless electric activity; PPV, positive predictive value; pVT, pulseless ventricular tachycardia; RF, random forest; RR, respiratory rate; RRS, Rapid Response System; SBP, systolic blood pressure; SHAP, shapley additive explanations; SPTTS, single-parameter-track-tigger-system; V.Fib, ventricular fibrillation; WSCH, Wonju severance Christian hospital.

AVAILABILITY OF DATA AND MATERIALS

This study included data from human subjects that potentially contained sensitive patient details. Owing to the legal and ethical constraints set by the participating entities and the Institutional Review Board of WSCH, these data cannot be publicly shared. To gain access to this data, please reach out to the Institutional Review Board of WSCH. Intellectual Property rights laws safeguard the code used in this study and restrict its public distribution. For inquiries regarding the code, please contact Lee Yeha, CEO of VUNO, at yeha.lee@vuno.co.

AUTHOR CONTRIBUTIONS

DY, YS, KJC, MC—designed the research study, DY, YS—wrote original draft. HY, YJK, JYP—collected initial data and was in charge of data acquisition. DY, YS, KJC—analyzed and interpreted the data. DY, YS—analyzed statistically. DY—supervised and reviewed the study. DY, MC—edited final manuscript. All authors read and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study protocol was reviewed and approved by the Ethics Committee and Institutional Review Board of WSCH in South Korea. The IRB number for this study was CR321150 and the approval date was 07 December 2021. The research undertaken was deemed to pose a minimal risk, and informed consent was not required because of the retrospective nature of the study. This study was conducted in accordance with the ethical standards of the Committee on Human Experimentation and the 1975 Declaration of Helsinki.

ACKNOWLEDGMENT

We would like to thank Editage (www.editage.com) and EnglishGo for English language editing.

FUNDING

This work was supported by the Korea Medical Device Development Fund grant funded by the Korean government (Ministry of Science and ICT, Ministry of Trade, Industry and Energy, Ministry of Health & Welfare, and Ministry of Food and Drug Safety) (Project Number: 1711195392, RS-2020-KD000030).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Sartini M, Carbone A, Demartini A, Giribone L, Oliva M, Spagnolo AM, *et al*. Overcrowding in emergency department: causes, consequences, and solutions-a narrative review. Healthcare. 2022; 10: 1625.

[2] Herring AA, Ginde AA, Fahimi J, Alter HJ, Maselli JH, Espinola JA, *et al*. Increasing critical care admissions from U.S. emergency departments, 2001–2009. Critical Care Medicine. 2013; 41: 1197–1204.

[3] Singer AJ, Thode Jr HC, Viccellio P, Pines JM. The association between length of emergency department boarding and mortality. Academic Emergency Medicine. 2011; 18: 1324–1329.

[4] Bhat R, Goyal M, Graf S, Bhooshan A, Teferra E, Dubin J, *et al*. Impact of post-intubation interventions on mortality in patients boarding in the emergency department. Western Journal of Emergency Medicine. 2014; 15: 708–711.

[5] Gunnerson KJ, Bassin BS, Havey RA, Haas NL, Sozener CB, Medlin RP, *et al*. Association of an emergency department-based intensive care unit with survival and inpatient intensive care unit admissions. JAMA Network Open. 2019; 2: e197584.

[6] Verma A, Vishen A, Haldar M, Jaiswal S, Ahuja R, Sheikh WR, *et al*. Increased length of stay of critically ill patients in the emergency department associated with higher in-hospital mortality. Indian Journal of Critical Care Medicine. 2021; 25: 1221–1225.

[7] Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA. A conceptual model of emergency department crowding. Annals of Emergency Medicine. 2003; 42: 173–180.

[8] McKenna P, Heslin SM, Viccellio P, Mallon WK, Hernandez C, Morley EJ. Emergency department and hospital crowding: causes, consequences, and cures. Clinical and Experimental Emergency Medicine. 2019; 6: 189–195.

[9] Savioli G, Ceresa IF, Guarnone R, Muzzi A, Novelli V, Ricevuti G, *et al*. Impact of coronavirus disease 2019 pandemic on crowding: a call to action for effective solutions to "access block". The Western Journal of Emergency Medicine. 2021; 22: 860–870.

[10] Pearce S, Marchand T, Shannon T, Ganshorn H, Lang E. Emergency department crowding: an overview of reviews describing measures causes, and harms. Internal and Emergency Medicine. 2023; 18: 1137–1158.

[11] Jones S, Moulton C, Swift S, Molyneux P, Black S, Mason N, *et al*. Association between delays to patient admission from the emergency department and all-cause 30-day mortality. Emergency Medicine Journal. 2022; 39: 168–173.

[12] Tej Prakash S, Brunda RL, Sakshi Y, Sanjeev B. Practice Changing Innovations for Emergency Care during the COVID-19 Pandemic in Resource Limited Settings. In: Vijay K, editor. SARS-CoV-2 Origin and COVID-19 Pandemic Across the Globe (pp. 195–208). IntechOpen: Rijeka. 2021.

[13] Tedesco D, Capodici A, Gribaudo G, Di Valerio Z, Montalti M, Salussolia A, *et al*. Innovative health technologies to improve emergency department performance. European Journal of Public Health. 2022; 32; 131–169.

[14] Weingart SD, Sherwin RL, Emlet LL, Tawil I, Mayglothling J, Rittenberger JC. ED intensivists and ED intensive care units. The American Journal of Emergency Medicine. 2013; 31: 617–620.

[15] Jeong H, Jung YS, Suh GJ, Kwon WY, Kim KS, Kim T, *et al*. Emergency physician-based intensive care unit for critically ill patients visiting emergency department. The American Journal of Emergency Medicine. 2020; 38: 2277–2282.

[16] Korean National Emergency Medical Center. Statistical yearbook of emergency medical service, 2021. 2021. Available at: `https://www.e-gen.or.kr/nemc/statistics_annual_report.do` (Accessed: 15 October 2023).

[17] Kim JH, Kim J, Bae S, Lee T, Ahn JJ, Kang BJ. Intensivists' direct management without residents may improve the survival rate compared to high-intensity intensivist staffing in academic intensive care units: retrospective and crossover study design. Journal of Korean Medical Science. 2020; 35: e19.

[18] Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, *et al*. Heart disease and stroke statistics-2023 update: a report from the american heart association. Circulation. 2023; 147: e93–e621.

[19] Nijor S, Rallis G, Lad N, Gokcen E. Patient safety issues from information overload in electronic medical records. Journal of Patient Safety. 2022; 18: e999–e1003.

[20] Cox EGM, Wiersema R, Eck RJ, Kaufmann T, Granholm A, Vaara ST, *et al*. External validation of mortality prediction models for critical illness reveals preserved discrimination but poor calibration. Critical Care Medicine. 2023; 51: 80–90.

[21] Le Gall JR. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA. 1993; 270: 2957–2963.

[22] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Critical Care Medicine. 1985; 13: 818–829.

[23] Kwon J, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. Journal of the American Heart Association. 2018; 7: e008678.

[24] Pandya S, Gadekallu TR, Reddy PK, Wang W, Alazab M. InfusedHeart: a novel knowledge-infused learning framework for diagnosis of cardiovascular events. IEEE Transactions on Computational Social Systems. 2022; 1–10.

[25] Liu S, Wang X, Xiang Y, Xu H, Wang H, Tang B. Multi-channel fusion LSTM for medical event prediction using EHRs. Journal of Biomedical Informatics. 2022; 127: 104011.

[26] Park SJ, Cho K, Kwon O, Park H, Lee Y, Shim WH, *et al*. Development and validation of a deep-learning-based pediatric early warning system: a single-center study. Biomedical Journal. 2022; 45: 155–168.

[27] Lee YJ, Cho K, 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. 2021; 163: 78–85.

[28] Lee Y, Kwon J, Lee Y, Park H, Cho H, Park J. Deep learning in the medical domain: predicting cardiac arrest using deep learning. Acute and Critical Care. 2018; 33: 117–120.

[29] Kang D, Cho K, Kwon O, Kwon J, Jeon K, Park H, *et al*. Artificial

intelligence algorithm to predict the need for critical care in prehospital emergency medical services. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine. 2020; 28: 17.

[30] Cho K, Kwon O, Kwon J, Lee Y, Park H, Jeon K, *et al*. Detecting patient deterioration using artificial intelligence in a rapid response system. Critical Care Medicine. 2020; 48: e285–e289.

[31] Chandriah KK, Naraganahalli RV. RNN/LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. Multimedia Tools and Applications. 2021; 80: 26145–26159.

[32] Jacobs I, Nadkarni V, Bahr J, Berg RA, Billi JE, Bossaert L, *et al*. Cardiac arrest and cardiopulmonary resuscitation outcome reports. Circulation. 2004; 110: 3385–3397.

[33] Song MJ, Lee YJ. Strategies for successful implementation and permanent maintenance of a rapid response system. The Korean Journal of Internal Medicine. 2021; 36: 1031–1039.

[34] Smith GB, Prytherch DR, Schmidt PE, Featherstone PI. Review and performance evaluation of aggregate weighted 'track and trigger' systems. Resuscitation. 2008; 77: 170–179.

[35] Cvach M. Monitor alarm fatigue: an integrative review. Biomedical Instrumentation & Technology. 2012; 46: 268–277.

[36] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the achilles heel of predictive analytics. BMC Medicine. 2019; 17: 230.

[37] Fang AHS, Lim WT, Balakrishnan T. Early warning score validation methodologies and performance metrics: a systematic review. BMC Medical Informatics and Decision Making. 2020; 20: 111.

[38] Xie Y, Chen J, Xu J, Shen B, Liao J, Teng J, *et al*. Early goal-directed renal replacement therapy in acute decompensated heart failure patients with cardiorenal syndrome. Blood Purification. 2022; 51: 251–259.

[39] Jessen MK, Vallentin MF, Holmberg MJ, Bolther M, Hansen FB, Holst JM, *et al*. Goal-directed haemodynamic therapy during general anaesthesia for noncardiac surgery: a systematic review and meta-analysis. British Journal of Anaesthesia. 2022; 128: 416–433.

[40] Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, *et al*. Early goal-directed therapy in the treatment of severe sepsis and septic shock. New England Journal of Medicine. 2001; 345: 1368–1377.

[41] Poncette AS, Spies C, Mosch L, Schieler M, Weber-Carstens S, Krampe H, *et al*. Clinical requirements of future patient monitoring in the intensive care unit: qualitative study. JMIR Medical Informatics. 2019; 7: e13064.

[42] Romare C, Anderberg P, Sanmartin Berglund J, Skär L. Burden of care related to monitoring patient vital signs during intensive care; a descriptive retrospective database study. Intensive and Critical Care Nursing. 2022; 71: 103213.

[43] Gasciauskaite G, Lunkiewicz J, Roche TR, Spahn DR, Nöthiger CB, Tscholl DW. Human-centered visualization technologies for patient monitoring are the future: a narrative review. Critical Care. 2023; 27: 254.

[44] Allyn J, Devineau M, Oliver M, Descombes G, Allou N, Ferdynus C. A descriptive study of routine laboratory testing in intensive care unit in nearly 140,000 patient stays. Scientific Reports. 2022; 12: 21526.

[45] Kim J, Chae M, Chang HJ, Kim YA, Park E. Predicting cardiac arrest and respiratory failure using feasible artificial intelligence with simple trajectories of patient data. Journal of Clinical Medicine. 2019; 8: 1336.

[46] Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC Medicine. 2015; 13: 1.

[47] Sung M, Hahn S, Han CH, Lee JM, Lee J, Yoo J, *et al*. Event prediction model considering time and input error using electronic medical records in the intensive care unit: retrospective study. JMIR Medical Informatics. 2021; 9: e26426.

[48] Kim J, Park YR, Lee JH, Lee JH, Kim YH, Huh JW. Development of a real-time risk prediction model for in-hospital cardiac arrest in critically Ill patients using deep learning: retrospective study. JMIR Medical Informatics. 2020; 8: e16349.

[49] 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. Computer Methods and Programs in Biomedicine. 2022; 214: 106568.

[50] Smith SK, Sincich T. An empirical analysis of the effect of length of forecast horizon on population forecast errors. Demography. 1991; 28: 261–274.

[51] Bounoua Z, Mechaqrane A. Hourly and sub-hourly ahead global horizontal solar irradiation forecasting *via* a novel deep learning approach: a case study. Sustainable Materials and Technologies. 2023; 36: e00599.

[52] Chandra R, Goyal S, Gupta R. Evaluation of deep learning models for multi-step ahead time series prediction. IEEE Access. 2021; 9: 83105–83123.

[53] Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the electronic medical record. Journal of Biomedical Informatics. 2013; 46: 837–848.

[54] Bell D, Baker J, Williams C, Bassin L. A trend-based early warning score can be implemented in a hospital electronic medical record to effectively predict inpatient deterioration. Critical Care Medicine. 2021; 49: e961–e967.

[55] Lee KJ, Tilling KM, Cornish RP, Little RJA, Bell ML, Goetghebeur E, *et al*. Framework for the treatment and reporting of missing data in observational studies: the treatment and reporting of missing data in observational studies framework. Journal of Clinical Epidemiology. 2021; 134: 79–88.

[56] Ghaferi AA, Schwartz TA, Pawlik TM. STROBE reporting guidelines for observational studies. JAMA Surgery. 2021; 156: 577–578.

[57] Combe B. SP0183 2016 update of the EULAR recommendations for the management of early arthritis. Annals of the Rheumatic Diseases. 2016; 75: 44–45.

[58] Mitchell OJL, Dewan M, Wolfe HA, Roberts KJ, Neefe S, Lighthall G, *et al*. Defining physiological decompensation: an expert consensus and retrospective outcome validation. Critical Care Explorations. 2022; 4: e0677.

[59] Hillman K, Chen J, Cretikos M, Bellomo R, Brown D, Doig G, F, *et al*. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. The Lancet. 2005; 365: 2091–2097.

[60] Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. 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. 2013; 84: 465–470.

[61] Pimentel MAF, Redfern OC, Malycha J, Meredith P, Prytherch D, Briggs J, *et al*. Detecting deteriorating patients in the hospital: development and validation of a novel scoring system. American Journal of Respiratory and Critical Care Medicine. 2021; 204: 44–52.

[62] Zimmerman JE, Kramer AA, Knaus WA. Changes in hospital mortality for United States intensive care unit admissions from 1988 to 2012. Critical Care. 2013; 17: R81.

[63] Capuzzo M, Volta CA, Tassinati T, Moreno RP, Valentin A, Guidet B, *et al*. Hospital mortality of adults admitted to intensive care units in hospitals with and without intermediate care units: a multicentre european cohort study. Critical Care. 2014; 18: 551.

[64] Halpern NA, Pastores SM. Critical care medicine beds, use, occupancy, and costs in the united states. Critical Care Medicine. 2015; 43: 2452–2459.

[65] Pronovost PJ, Needham DM, Waters H, Birkmeyer CM, Calinawan JR, Birkmeyer JD, *et al*. Intensive care unit physician staffing: financial modeling of the Leapfrog standard. Critical Care Medicine. 2004; 32: 1247–1253.

[66] Halpern NA, Tan KS, DeWitt M, Pastores SM. Intensivists in U.S. acute care hospitals. Critical Care Medicine. 2019; 47: 517–525.

[67] Jatoi NN, Awan S, Abbasi M, Marufi MM, Ahmed M, Memon SF, *et al*. Intensivist and COVID-19 in the United States of America: a narrative review of clinical roles, current workforce, and future direction. The Pan African Medical Journal. 2022; 41: 210.

How to cite this article: Yunseob Shin, Kyung-jae Cho, Mineok Chang, Hyun Youk, Yoon Ji Kim, Ji Yeong Park, *et al*. The development and validation of a novel deep-learning algorithm to predict in-hospital cardiac arrest in ED-ICU (emergency department-based intensive care units): a single center retrospective cohort study. Signa Vitae. 2024; 20(4): 83-98. doi: 10.22514/sv.2024.045.