mathematics

MDPI

Article

Machine Learning-Based Cardiac Arrest Prediction for Early Warning System

Minsu Chae [1], Hyo-Wook Gil [2], Nam-Jun Cho [2] and Hwamin Lee [1,*]

[1] Department of Medical Informatics, College of Medicine, Korea University, Seoul 02841, Korea; minsuchae@korea.ac.kr
[2] Department of Internal Medicine, Soonchunhyang University Cheonan Hospital, Cheonan 31151, Korea; hwgil@schmc.ac.kr (H.-W.G.); chonj@schmc.ac.kr (N.-J.C.)
* Correspondence: hwamin@korea.ac.kr; Tel.: +82-2-2286-1125

Abstract: The early warning system detects early and responds quickly to emergencies in high-risk patients, such as cardiac arrest in hospitalized patients. However, traditional early warning systems have the problem of frequent false alarms due to low positive predictive value and sensitivity. We conducted early prediction research on cardiac arrest using time-series data such as biosignal and laboratory data. To derive the data attributes that affect the occurrence of cardiac arrest, we performed a correlation analysis between the occurrence of cardiac arrest and the biosignal data and laboratory data. To improve the positive predictive value and sensitivity of early cardiac arrest prediction, we evaluated the performance according to the length of the time series of measured biosignal data, laboratory data, and patient data range. We propose a machine learning and deep learning algorithm: the decision tree, random forest, logistic regression, long short-term memory (LSTM), gated recurrent unit (GRU) model, and the LSTM–GRU hybrid model. We evaluated cardiac arrest prediction models. In the case of our proposed LSTM model, the positive predictive value was 85.92% and the sensitivity was 89.70%.

Keywords: cardiac arrest; machine learning; deep learning; early warning system

MSC: 94A16; 68T07; 4008

Citation: Chae, M.; Gil, H.-W.; Cho, N.-J.; Lee, H. Machine Learning-Based Cardiac Arrest Prediction for Early Warning System. *Mathematics* **2022**, *10*, 2049. https://doi.org/10.3390/math10122049

Academic Editors: Heui Seok Lim, Sanghyuk Lee, Yeongwook Yang, Imatitikua Aiyanyo and Ioannis G. Tsoulos

Received: 2 May 2022
Accepted: 9 June 2022
Published: 13 June 2022

1. Introduction

Cardiac arrest is a disease in which the heart stops. In the U.S, 356,000 people experience cardiac arrest every year [1]. In the Republic of Korea, 30,000 people experience cardiac arrest every year [2]. In cardiac arrest, the golden time is less than 3 min [3]. Delays in cardiopulmonary resuscitation lead to death. Therefore, cardiac arrest is important for early prediction. Cardiac arrest has been studied as having a precursor symptom or asymptomatic precursor symptoms [4,5]. To determine a patient's health, hospitals measure biosignal data and laboratory data via medical sensors and blood. The hospital operated a rapid response team (RRT) to manage ill patients such as cardiac arrest patients. RRT uses an early warning system (EWS) such as the national early warning score (NEWS) or the modified early warning score (MEWS). However, EWS has a low positive predictive value (PPV) and false alarms [6].

Recently, machine learning has been applied in healthcare [7–11]. Additionally, machine learning has been applied in the early prediction of cardiac arrest [6,12–18]. Sbrollini et al. [19] and Ibrahim et al. [20] developed models for predicting myocardial infarction, which is a precursor to cardiac arrest, using electrocardiograms (ECGs). However, it is difficult make the early prediction of cardiac arrest based on myocardial infarction because cardiac arrest is not necessarily preceded by myocardial infarction. Yosuf El Saadany et al. developed a wireless Internet of Things (IoT) device that predicts cardiac

arrest based on abnormal patterns on measured ECG [21]. However, the ECG has a disadvantage in that the measurement equipment requires being worn by each patient. They studied the early prediction of cardiac arrest based on machine learning such as decision tree, random forest, logistic regression, support vector machine, and recurrent neural network [6,13–18]. The PPV of the prediction cardiac arrest algorithm improves the traditional EWS, but the PPV of the prediction cardiac arrest algorithm is lower than 10% [6]. Thus, they have a problem with false alarms. Early cardiac arrest prediction mainly used age, sex, race, biosignal, and laboratory data. Churpek et al. proposed maximum respiration rate and minimum diastolic blood pressure (DBP) as important parameters in predicting cardiac arrest [22]. In this research, we developed an early prediction model for cardiac arrest that improved the PPV and the sensitivity of traditional EWS based on shallow and deep learning.

2. Materials

We performed a retrospective cohort study at Soonchunhyang University Cheonan Hospital in the Republic of Korea. This research population consisted of patients admitted to Soonchunhyang University Cheonan Hospital between January 2016 and June 2019. This research excludes patients under 18 years of age and patients who were dead or with cardiac arrest within 8 h of admission. In addition, we excluded patients whose albumin, bilirubin, creatinine, platelet, hemoglobin, or white blood cell rates were never measured. In our previous research [23], we used input parameters such as laboratory data and laboratory check variables, but it was not properly considered. Table 1 shows the characteristics of our study population.

Table 1. Characteristics of the study population.

Characteristics	Description
Study period	January 2016–June 2019
Total patients, n	34,452
Patients with in-hospital cardiac arrest, n	573
Number of features, n	14
Number of data for each patient, n	72
Sequence data slice size	8
Age, years, (mean ± SD)	58.6 ± 17.0
Males, n (%)	16,760 (48.7%)
Hospital	Soonchunhyang University Cheonan Hospital

SD: standard deviation.

Medical sensors measured the biosignal data and laboratory data. The biosignal data were manually typed into the hospital information system. The biosignal data therefore suffered from human error. We used electronic health records (EHRs) data. Table 2 shows the EHRs parameters. We ignored the abnormal biosignal data and biosignal data with human error.

Table 2. Electronic health records (EHR) data parameters.

Variable	Description
Age	Age at hospitalization
Sex	Male (1) or female (2)
DBP	Diastolic blood pressure ($30 \leq$ DBP ≤ 300, mmHg)

Table 2. *Cont.*

Variable	Description
SBP	Systolic blood pressure (30 ≤ SBP ≤ 300, mmHg)
Body temperature	Body temperature (30 ≤ body temperature ≤ 45)
Respiratory rate	Breaths per minute (3 ≤ breath ≤ 60)
Blood pressure	Blood pressure (30 ≤ blood pressure ≤ 300, mmHg)
Albumin	Albumin values (laboratory data)
Bilirubin	Bilirubin values (laboratory data)
Creatinine	Creatinine values (laboratory data)
PLT	Number of platelets (laboratory data)
Hb	Number of hemoglobin (laboratory data)
WBC	Number of white blood cells (laboratory data)
AST	Aspartate aminotransferase values (laboratory data)
ALT	Alanine aminotransferase values (laboratory data)

DBP: diastolic blood pressure; SBP: systolic blood pressure; PLT: platelets; Hb: hemoglobin; WBC: white blood cell; AST: aspartate aminotransferase; ALT: alanine aminotransferase.

3. Methods

We used the TensorFlow, Keras, and scikit-learn libraries for machine learning [24–26]. We expressed data in three dimensions using a recurrent neural network (RNN) model. Figure 1 shows the workflow of cardiac arrest early prediction.

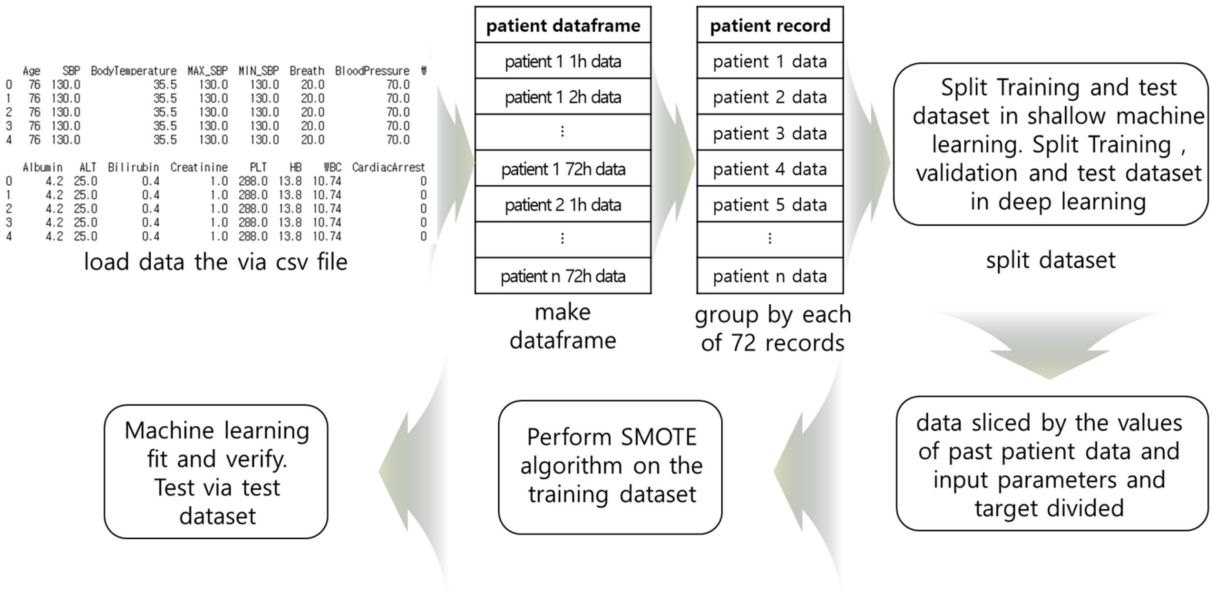

Figure 1. The workflow of cardiac arrest prediction.

3.1. Shallow and Deep Learning

3.1.1. Decision Tree

Decision Tree is a supervised learning method used for classification and regression. It consists of nodes and branches. When it is constructed by recursively evaluation different features and using at each node the feature that best splits the data. Table 3 shows variables of the decision tree. At each node a variable is evaluated to decide which path to follow and chooses the left leaf node or right leaf node based on a threshold. It is a similar rule-based expert system. However, it is trained via a dataset. It uses a heuristic algorithm such as information gain [27]. The decision tree computes the maximum impurity considering

a feature and a threshold. In this paper, the decision tree selects one feature from the biosignal sign and laboratory data. The decision tree calculates the maximum impurity to classify cardiac arrest patients and patients without cardiac arrest.

Table 3. Variables of the decision tree.

Variable	Description
p	The function of calculating classes ratio.
count	The function of count label data. In this paper, the function counts cardiac arrest patients and patients without cardiac arrest.
dataset	Input element and target dataset. The EHRs dataset and event of cardiac arrest in this paper.
n	The number of the dataset.
getLeftNode	The function returns the dataset in which the value of the feature is less or equal to the threshold.
getRightNode	The function returns the dataset in which the value of the feature is greater than the threshold.

The decision tree calculates the impurity through Equation (1). The impurity is used to calculate the distribution of cardiac arrest patients and patients without cardiac arrest.

$$I(dataset) = 1 - \ p(count(dataset,\ 0))^2 - p(count(dataset,\ 1))^2 \tag{1}$$

The maximum information gain is calculated through Equation (2), and the optimal feature and threshold are obtained through iterative calculations.

$$\begin{aligned} IG(dataset,\ feature,\ threshold) = I(dataset) \\ - \tfrac{count(0)}{n} I(getLeftNode(dataset,\ feature,\ threshold)) \\ - \tfrac{count(1)}{n} I(getRightNode(dataset,\ feature,\ threshold)) \end{aligned} \tag{2}$$

3.1.2. Random Forest

Random Forest is an ensemble algorithm-based decision tree [27,28]. The random forest consists of many decision trees and uses bootstrapping [28]. It predicts based on the voting of many decision trees. Random forest bootstraps via the training data of EHRs data and predicts cardiac arrest events using voting by many decision trees.

3.1.3. Logistic Regression

The logistic regression algorithm is a linear classifier algorithm [27]. The logistic regression algorithm applies a logistic function to the result of a linear classifier [27,29]. A logistic regression calculated gradient and bias to classify cardiac arrest patients based on the EHRs data.

3.1.4. Recurrent Neural Network

The recurrent neural network (RNN) model is fit for time-series data. The RNN model calculates the hidden state and considers the result of previous hidden states and the current input parameters [30]. The vanilla RNN cell considers short-term memory. Table 4 shows the variable used in the RNN cell [30].

Vanilla RNN cells are calculated by sequentially calculating the previous hidden state and time series data as shown in Equation (3) [30]. The RNN model trains to make the early prediction of cardiac arrest by calculating weights using EHRs data measured at one-hour intervals.

$$H_t = tanh(h_{t-1}W_h + W_x x_t + b) \tag{3}$$

Table 4. Variables of the RNN cell.

Variable	Description
tanh	Hyperbolic tangent function.
h	Hidden state. h_t means a hidden state of the t time.
x	Input value. x_t means the input value of the t time.
W	Weight matrices. W_h means the weight of the hidden state. W_x means the weight of the input value.
b	The bias of the hidden value.

The long short-term memory (LSTM) cell solves the long-term dependency problem proposed by Horchreiter et al. [31]. Table 5 shows the variable used in the LSTM cell [31].

Table 5. Variables of the LSTM cell.

Variable	Description
σ	Logistic sigmoid function.
i	Value of input gate. I_t means the value of the input gate of the t time.
f	Value of forget gate. f_t means the value of the forget gate of the t time.
o	Value of output gate. o_t means the value of the output gate of the t time.
c	Long-term memory. C_t means the long-term memory of the t time.
h	Hidden state. H_t means a hidden state of the t time.
W	Weight matrices of matrices. $W_{\{x \text{ or } h\}\{i \text{ or } f \text{ or } o \text{ or } c\}}$ means the weight of the input value or hidden state and input gate or forget gate or output gate or long-term memory.
b	The bias of hidden value. b_i means the bias of input gate. b_f means the bias of forget gate. b_o means the bias of output gate. b_c means the bias of long-term memory.

In this paper, the importance weights of cardiac arrest are calculated in sequence through Equation (4) [31].

$$i_t = \sigma(W_{\text{xii}} x_t + W_{hi} h_{t-1} + b_i) \tag{4}$$

In this paper, LSTM cells calculate unnecessary weights for predicting cardiac arrest through Equation (5), and long-term weights are updated through Equation (6) [31].

$$f_t = \sigma\left(W_{xf} x_t + W_{hf} h_{t-1} + b_f\right) \tag{5}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \tag{6}$$

In this paper, the LSTM cells calculate the output result of the LSTM cell through Equation (7), and make the early prediction of cardiac arrest within 8 h through Equation (8) [31].

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o) \tag{7}$$

$$h_t = o_t \cdot tanh(c_t) \tag{8}$$

We used Adam algorithm as the optimizer of the LSTM model. The Adam algorithm is one of the stochastic gradient descent methods that considers adaptive estimation and moments [32]. The LSTM model requires input features and previous weights to calculate weights. In this paper, input features are EHRs data transmitted to the input gate, the forget gate, and the output gate. The input gate calculates the weight of cardiac arrest prediction through Equation (4) and transmits it to the forget gate. The forget gate calculates the unnecessary weights for predicting cardiac arrest through Equation (5). It removes unnecessary weights from the long-term memory and updates the weights through Equation (6). The updated long-term memory is transferred to the output gate. The output gate calculates the weights to predict cardiac arrest through Equations (7) and (8) based on short-term

memory and long-term memory and EHRs data and stores the short-term memory. We used the Adam algorithm to calculate the weight through backpropagation. We organized an output layer after the LSTM model to predict cardiac arrest within 8 h.

Our LSTM models are subsequently constructed of the LSTM layer, dropout layer, LSTM layer, dropout layer, LSTM layer, dropout layer, LSTM layer, dropout layer, and dense layer.

The gated recurrent unit (GRU) cell designed by Cho et al. [33] improved the processing time compared with the LSTM cell and obtained a result similar to the LSTM cell. Table 6 shows the variable used in the GRU cell [33].

Table 6. Variables of the GRU cell.

Variable	Description
σ	Logistic sigmoid function.
$tanh$	Hyperbolic tangent function.
r	Value of reset gate. r_t means the value of the reset gate of the t time.
z	Value of update gate. z_t means the value of the update gate of the t time.
h	Hidden state. h_t means a hidden state of the t time.
W	Weight matrices of matrices. $W_{\{x \text{ or } h\}\{r \text{ or } z\}}$ means the weight of input value or hidden state and reset gate or update gate.
B	The bias of hidden value. B_r means bias of reset gate. B_z means bias of update gate.

In the GRU cell, it is decided whether to use the previous result weight or update the weight of the cardiac arrest prediction to reduce the number of operations. In this paper, the GRU cell determines whether to use the previous weight through Equation (9) [33], and calculates the cardiac arrest weight using the EHR data.

$$R_t = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{9}$$

In this paper, the GRU cell determines whether to use the previous cardiac arrest weight or calculate the cardiac weight using EHRs data through Equation (10) [33].

$$Z_t = \sigma(W_{xz}x_t + W_{hz}h_{t-1} + b_z) \tag{10}$$

In this paper, the GRU cell predicts cardiac arrest using EHRs data through Equation (11) [33].

$$H_t = (1 - z_t) \cdot tanh(W_{hr}(r_t \cdot h_{t-1}) + W_{hz}h_{t-1} + b_r) + z_t \cdot h_{t-1} \tag{11}$$

The GRU model requires input features and previous weights to calculate weights. In this paper, input features are EHRs data transmitted to the reset gate and the update gate. The reset gate calculates the weight of cardiac arrest prediction through Equation (9). The update gate calculates whether to use the stored weight or the newly calculated weight through Equation (10). The short-term memory determines whether to use the previous weight or to store the newly calculated weight through Equation (11). We used the Adam algorithms to calculate weight through backpropagation. We organized an output layer after the GRU model to predict cardiac arrest within 8 h.

Our GRU model consists of a stack of the GRU layer, dropout layer, GRU layer, dropout layer, GRU layer, dropout layer, GRU layer, dropout layer, and dense layer. Our LSTM–GRU hybrid model is constructed the following LSTM layer, dropout layer, LSTM layer, dropout layer, GRU layer, dropout layer, GRU layer, dropout layer, and dense layer.

Our model hyperparameters are as follows: the activation function of the LSTM and GRU layers is hyperbolic tangent; the activation function of the dense layer is sigmoid; the

maximum epoch is 100; the monitoring value of early stop is the validation f1 score; the loss function is binary cross-entropy; and the optimizer is Adam.

3.2. Synthetic Minority Oversampling Technique

Synthetic minority oversampling technique (SMOTE) is a kind of oversampling algorithm proposed by Chawlas et al. [34]. In this research, cardiac arrest patients and other patients are unbalanced. Therefore, it is difficult for machine learning to correctly predict cardiac arrest. We solved the overfitting problem by applying the SMOTE algorithm to the training dataset. We applied the SMOTE algorithm with a majority to minority ratio of 1:1.

3.3. Data Preprocessing

Time series data require equal measurement intervals. The medical staff measures biosignal and laboratory data in consideration of the patients' health conditions. Therefore, the measurement interval is different for each patient. In the case of medical staff typing a biosignal, it may have human error. Therefore, EHRs data require data preprocessing. We changed the measurement interval to one hour because the measurement interval is one hour in ICU patients. The EHRs data express the patient information, measurement item, and measurement value. Patient information includes patient identification (ID), age, and sex. Hospitals measure biosignal data and laboratory data when monitoring patient health; therefore, the measurement intervals of each patient may be different. We replaced the missing value with the last measured values. Each patient data range is different. We only used 72 h of patient data.

Machine learning is generally split into training and test datasets or training and validation and test datasets. In our patient data, if the data of the same patient is divided into a training dataset, a validation dataset, and a test dataset, this may cause an overfitting problem. Considering this overfitting problem, we grouped the same patient data into one group and then split the dataset.

We split the training and test datasets based on the grouped patients and shuffled the datasets via the train_test_split method provided by scikit-learn [26]. The ratio of the training to test datasets was 9:1. Additionally, we split the training and validation datasets and shuffled the datasets via the train_test_split method based on the training dataset in deep learning. The ratio of the training to validation datasets was 9:1. We data sliced the data of the grouped patients into past patient data. We applied the SMOTE algorithm to the training dataset.

4. Result

4.1. Performance Evaluation Method

In general, machine learning uses accuracy as a performance evaluation for classification. In healthcare, machine learning considers PPV and sensitivity. Figure 2 shows four types of data prediction.

		Actual Cardiac arrest within 8 hours	
		Cardiac Arrest	**Non-Cardiac Arrest**
Result of Prediction	**Cardiac Arrest**	True Positive (TP) Detect patients of cardiac arrest	False Positive (FP) Has False alarms
	Non-Cardiac Arrest	False Negative (FN) Does not detect patient with cardiac arrest	True Negative (TN) Detects patients without cardiac arrest

Figure 2. Four types of data prediction.

PPV was calculated by Equation (12). PPV considers the predicted cardiac arrest. Sensitivity considers cardiac arrest patients. The F1 score considers both PPV and sensitivity with the same weight.

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

A negative predictive value (NPV) was calculated by Equation (13). NPV considers the predicted non-cardiac arrest.

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \tag{13}$$

Sensitivity was calculated by Equation (14).

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

F1 score was calculated by Equation (15).

$$\text{F1 score} = 2 \times \frac{\text{PPV} \times \text{sensitivity}}{\text{PPV} + \text{sensitivity}} \tag{15}$$

4.2. Performance Evaluation Results according to Input Parameters

We evaluated according to the input parameters for the correlation analysis of the input parameters. The training dataset used 49–72 h from each patient, and the validation and test datasets used 1–72 h of data from each patient. We performed cardiac arrest predictions within 8 h based on 24 h of data. We set 24 h of data because this decreases the learning time. We compared each laboratory datum based on patient information (e.g., age, sex) and each biosignal datum (e.g., SBP, DBP, body temperature, breaths per minute, blood pressure) also based on patient information. We compared the maximum SBP and minimum SBP based on patient information and biosignal data. Table 7 shows the average PPV, sensitivity, and F1 score based on the LSTM model via input parameters. Each parameter was performed five times.

Table 7. Performance evaluation of each parameter.

Input Parameters	PPV	NPV	Sensitivity	F1 Score
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure	3.07%	99.89%	61.27%	0.0583
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, albumin	4.06%	99.94%	78.68%	0.0771
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, bilirubin	3.38%	99.91%	69.30%	0.0643
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, creatinine	2.80%	99.9%	67.63%	0.0533
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, PLT	6.85%	99.91%	69.08%	0.1150
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, Hb	3.82%	99.91%	67.85%	0.0722
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, WBC	3.36%	99.9%	64.56%	0.0639
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, AST	2.80%	99.92%	72.15%	0.0536

Table 7. *Cont.*

Input Parameters	PPV	NPV	Sensitivity	F1 Score
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, ALT	2.35%	99.94%	79.04%	0.0454
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, AST, ALT	2.38%	99.93%	76.40%	0.0460
Age, sex, SBP, body temperature, breaths per minute, blood pressure	1.98%	99.94%	79.04%	0.0385
Age, sex, DBP, body temperature, breaths per minute, blood pressure	15.77%	98.64%	51.14%	0.0519
Age, sex, body temperature, breath, blood pressure	2.36%	99.92%	73.73%	0.0519
Age, DBP, SBP, body temperature, breaths per minute, blood press	2.11%	99.93%	76.40%	0.0490
Sex, DBP, SBP, body temperature, breaths per minute, blood pressure	1.44%	99.91%	75.79%	0.0282
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, maximum SBP in 72 h	3.49%	99.92%	73.11%	0.0660
Age, sex, DBP, SBP, body temperature, breaths per minute, blood pressure, maximum SBP in 24 h	3.13%	99.92%	71.11%	0.0597
Age, DBP, SBP, body temperature, breaths per minute, blood pressure, minimum SBP in 72 h	3.68%	99.93%	77.63%	0.0696
Age, DBP, SBP, body temperature, breaths per minute, blood pressure, minimum SBP in 24 h	3.99%	99.89%	61.97%	0.0750
Age, DBP, SBP, body temperature, breaths per minute, blood pressure, maximum SBP in 72 h, minimum SBP in 72 h	5.23%	99.94%	81.18%	0.0976

PPV: positive predictive value; NPV: negative predictive value; DBP: diastolic blood pressure; SBP: systolic blood pressure.

We confirmed that the PPV or sensitivity increased each laboratory data based on patient information and biosignal data. We compared each AST, ALT, and both AST and ALT. In the case of AST, the average PPV was 0.5% higher than others. In the case of ALT, the average sensitivity was 79.04% higher than others. We confirmed that if sex and DBP were ignored, then the sensitivity was increased. We compared the maximum SBP in 72 h and minimum SBP in 72 h based on patient information—biosignal data—and the sensitivity was increased. Therefore, we used age, SBP, maximum SBP in 72 h, minimum SBP in 72 h, body temperature, breaths per minute, blood pressure, albumin, bilirubin, creatinine, PLT, Hb, WBC, ALT. Figure 3 shows the receiver operating characteristic (ROC) curve of each parameter.

4.3. Performance Evaluation Results according to the Number of Past Patient Data and Data Range

We evaluated according to the number of past patient data and data range. The training dataset used each data range, and the validation and test dataset used 1–72 h of data from each patient. Each LSTM model was performed five times. Table 8 shows the average PPV, sensitivity, and F1 score based on the LSTM model considering the past data and data range.

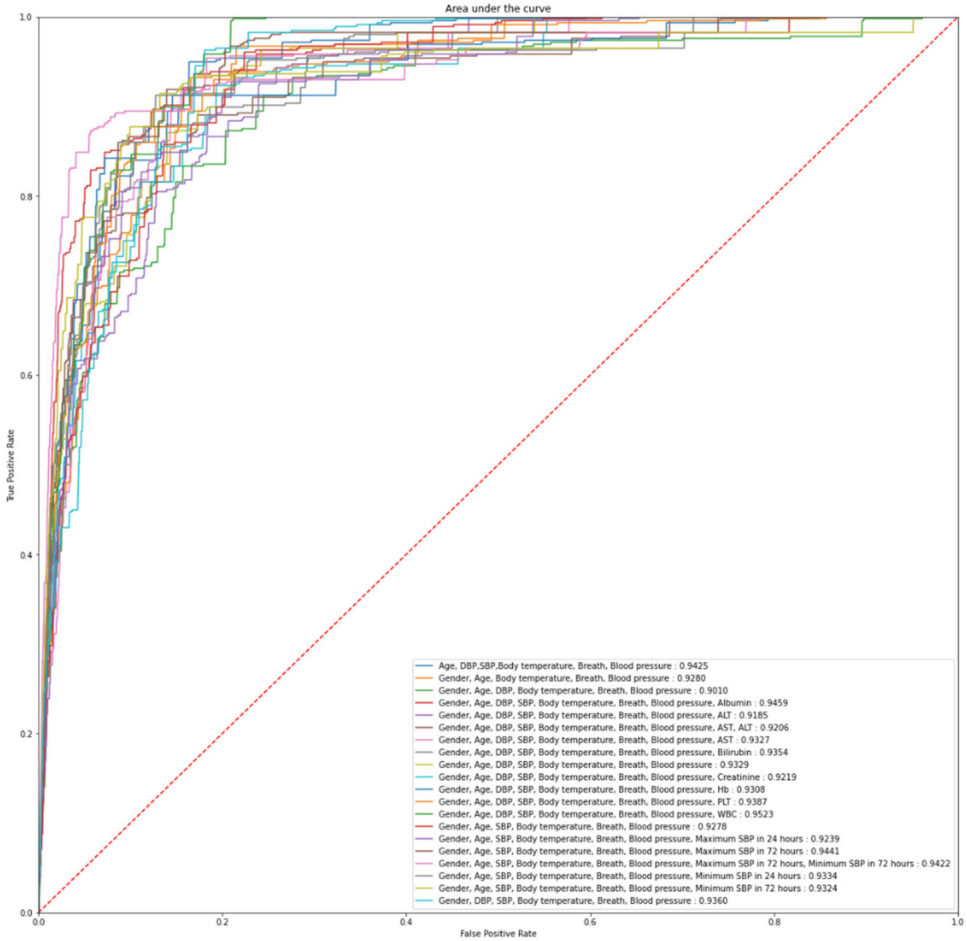

Figure 3. Receiver operating characteristic (ROC) curve of each parameter.

Table 8. Performance evaluation of the number of past patient data and patient data range.

Number of Past Patient Data	Patient Data Range	PPV	NPV	Sensitivity	F1 Score
8 h	1~72 h	9.03%	99.96%	81.23%	0.1623
	9~72 h	8.11%	99.97%	84.82%	0.1478
	17~72 h	9.04%	99.96%	81.53%	0.1626
	25~72 h	8.72%	99.97%	85.18%	0.1581
	33~72 h	8.25%	99.96%	79.61%	0.1493
	41~72 h	8.19%	99.96%	82.54%	0.1488
	49~72 h	7.22%	99.96%	80.53%	0.1321
	57~72 h	6.01%	99.96%	81.18%	0.1117
16 h	1~72 h	8.59%	99.96%	82.81%	0.1554
	9~72 h	9.29%	99.95%	80.18%	0.1664
	17~72 h	9.03%	99.95%	82.41%	0.1627
	25~72 h	8.63%	99.96%	81.75%	0.1558
	33~72 h	8.68%	99.96%	85.31%	0.1574
	41~72 h	9.74%	99.96%	82.85%	0.1742
	49~72 h	6.87%	99.96%	83.90%	0.1267

Table 8. *Cont.*

Number of Past Patient Data	Patient Data Range	PPV	NPV	Sensitivity	F1 Score
24 h	1~72 h	10.58%	99.97%	87.83%	0.1882
	9~72 h	10.83%	99.96%	84.69%	0.1919
	17~72 h	7.84%	99.96%	84.61%	0.1423
	25~72 h	10.65%	99.95%	83.64%	0.1889
	33~72 h	9.37%	99.96%	84.12%	0.1678
	41~72 h	9.47%	99.95%	80.95%	0.1690
32 h	1~72 h	15.71%	99.94%	83.16%	0.2641
	9~72 h	13.86%	99.94%	82.32%	0.2369
	17~72 h	10.31%	99.82%	81.49%	0.1822
	25~72 h	10.44%	99.93%	80.09%	0.1844
	33~72 h	9.45%	99.94%	83.16%	0.1687
40 h	1~72 h	19.39%	99.95%	87.41%	0.3173
	9~72 h	17.41%	99.94%	85.57%	0.2888
	17~72 h	11.63%	99.93%	82.59%	0.2020
	25~72 h	9.99%	99.94%	85.26%	0.1778
48 h	1~72 h	23.85%	99.93%	87.32%	0.3740
	9~72 h	17.05%	99.9%	82.63%	0.2802
	17~72 h	14.93%	99.91%	83.73%	0.2504
56 h	1~72 h	37.60%	99.88%	85.83%	0.5229
	9~72 h	30.59%	99.88%	85.75%	0.4438
64 h	1~72 h	61.26%	99.72%	83.60%	0.7001

PPV: positive predictive value; NPV: negative predictive value.

The dataset of this research has a few cases of cardiac arrest data. In the case of the number of past patient data of 64 h, PPV and sensitivity were highest. Figure 4 shows the receiver operating characteristic (ROC) curve of the past patient data range.

4.4. Performance Evaluation Result of Machine Learning Algorithm

We performed machine learning based on early cardiac arrest predictions. We changed the stratified k-fold for shallow machine learning and the unit size of layers for deep learning. We performed machine learning algorithms such as decision tree, random forest, logistic regression, LSTM model, GRU model, and LSTM–GRU hybrid model. Each algorithm was performed five times. Table 9 shows the average PPV, sensitivity, and F1 score based on each algorithm.

In the case of shallow machine learning, we used the stratified K-Fold algorithm provided by scikit-learn [26]. In the case of deep learning, we used EarlyStopping provided by TensorFlow [24] and its monitoring value was a validation F1 score. However, a few deep learning models overfit. We considered the maximum PPV of each algorithm.

The decision tree had low PPV and sensitivity. The random forest had the highest PPV but low sensitivity. Logistic regression had low PPV but had the highest sensitivity in shallow machine learning. Deep learning models were similar to the sensitivity of the logistic regression and had higher PPV than the logistic regression. Figure 5 shows the receiver operating characteristic (ROC) curve for each algorithm that has the best performance.

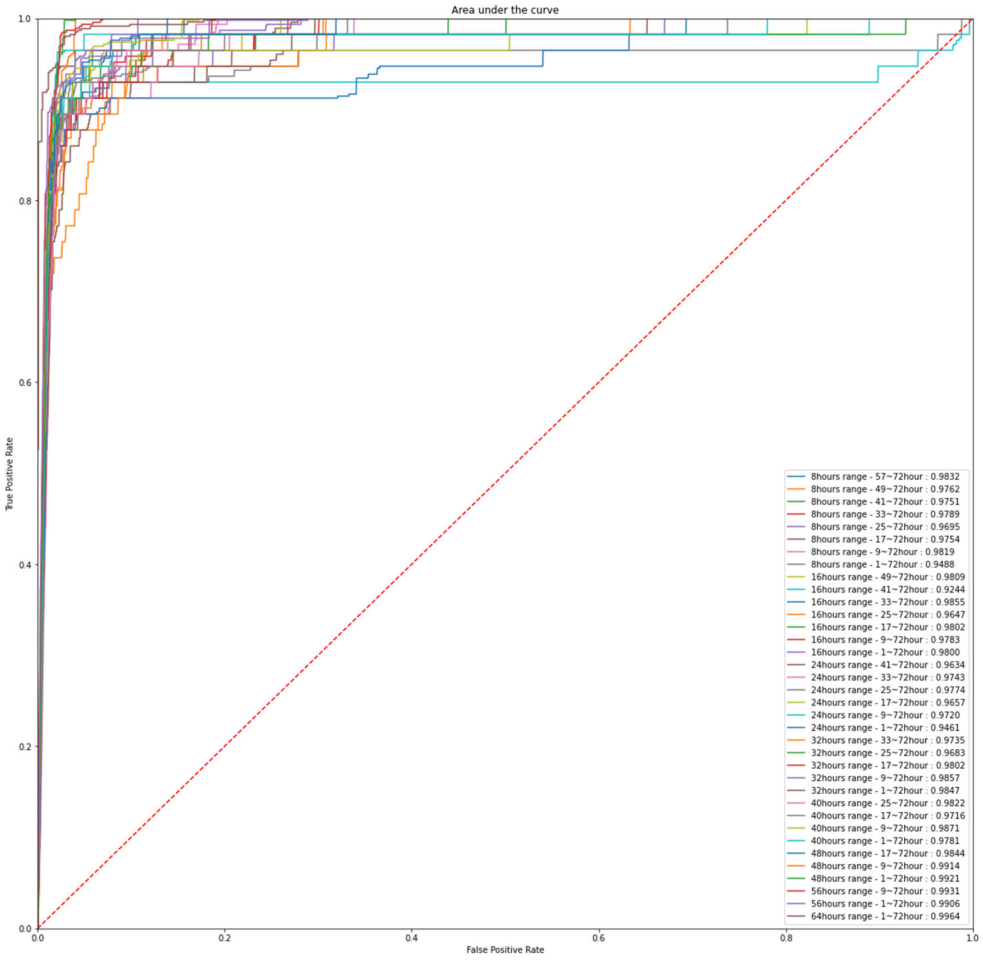

Figure 4. Receiver operating characteristic (ROC) curve of the past patient data range.

Table 9. Performance evaluation of the machine learning algorithm.

Algorithms	Hyperparameter	PPV	NPV	Sensitivity	F1 Score
Decision tree	K set 4	56.14%	99.21%	53.25%	0.5458
	K set 5	59.90%	99.26%	55.83%	0.5778
	K set 10	69.34%	99.29%	57.76%	0.6298
Random forest	K set 4	87.91%	99.22%	53.07%	0.6617
	K set 5	86.25%	99.26%	55.70%	0.6768
	K set 10	90.71%	99.26%	55.75%	0.6901
Logistic regression	K set 4	23.50%	99.80%	88.82%	0.3717
	K set 5	21.70%	99.81%	89.47%	0.3492
	K set 10	21.85%	99.84%	90.83%	0.3523
LSTM model	The unit size of layers set 32	72.20%	99.78%	87.15%	0.7883
	The unit size of layers set 64	73.95%	99.81%	89.04%	0.8075
	The unit size of layers set 96	71.68%	99.78%	87.15%	0.7756
	The unit size of layers set 128	67.85%	99.81%	88.64%	0.7583

Table 9. *Cont.*

Algorithms	Hyperparameter	PPV	NPV	Sensitivity	F1 Score
GRU model	The unit size of layers set 32	66.68%	99.78%	87.02%	0.7487
	The unit size of layers set 64	70.09%	99.79%	87.32%	0.7707
	The unit size of layers set 96	81.19%	99.85%	91.10%	0.8582
	The unit size of layers set 128	73.86%	99.86%	91.61%	0.8121
LSTM–GRU hybrid model	The unit size of layers set 32	63.78%	99.75%	84.96%	0.7233
	The unit size of layers set 64	61.05%	99.78%	86.93%	0.7000
	The unit size of layers set 96	60.43%	99.64%	79.05%	0.6692
	The unit size of layers set 128	69.35%	99.84%	90.35%	0.7825

PPV: positive predictive value; NPV: negative predictive value.

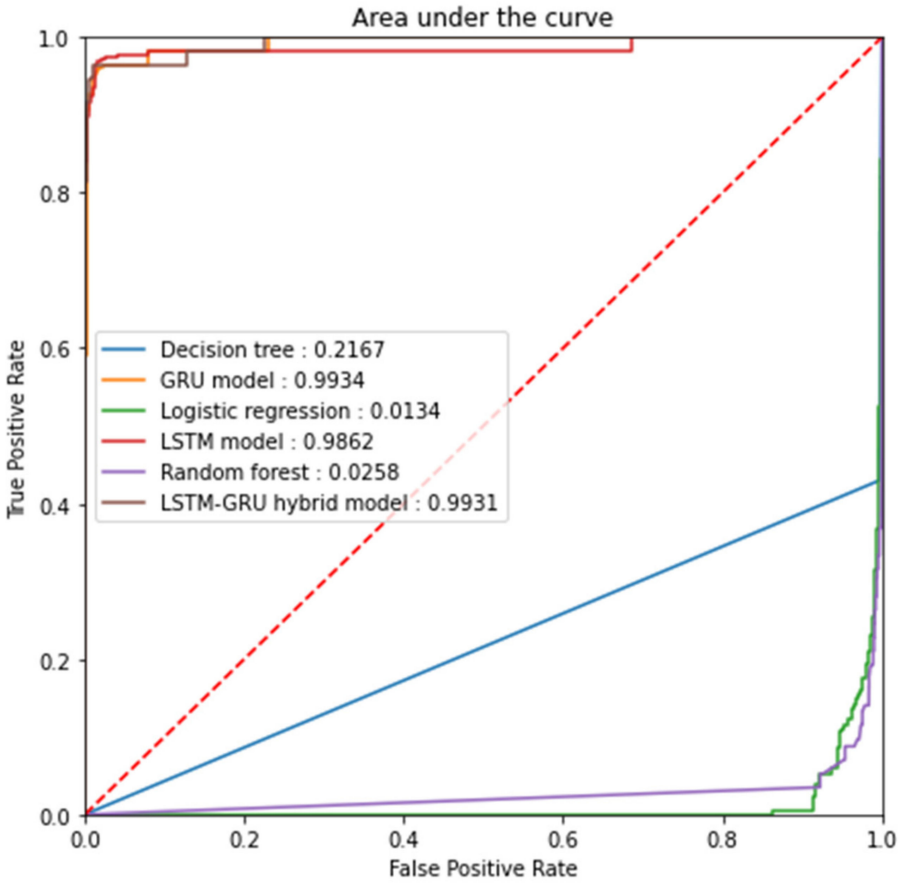

Figure 5. Receiver operating characteristic (ROC) curve of each algorithm.

5. Discussion

We proposed a cardiac arrest early prediction model to improve traditional EWS for PPV. We used biosignal data and laboratory data. We compared other cardiac arrest prediction algorithms based on biosignal and laboratory data. However, the performance of each algorithm is hard to evaluate because each algorithm uses different datasets. Table 10 shows the performance of the EWS and the proposed methods in this paper. Our proposed LSTM model had the highest PPV and F1 scores.

Table 10. Comparison of performance with previous studies for predicting in-hospital cardiac arrest.

Algorithms		PPV	NPV	Sensitivity	Specificity	F1 Score
Traditional EWS [6]	SPPTS	0.4%	99.9%	60.7%	77.0%	0.8
	MEWS ≥ 3	0.5%	99.9%	63.0%	79.9%	0.0100
	MEWS ≥ 4	0.6%	99.9%	49.3%	86.8%	0.0120
	MEWS ≥ 5	0.6%	99.9%	37.3%	90.6%	0.0130
	RF	0.4%	99.9%	75.3%	69.9%	0.0080
	LR	0.2%	99.9%	76.3%	34.6%	0.0040
Joon-myoung Kwon et al. [6]	DEWS ≥ 2.9	0.5%	99.9%	75.7%	76.5%	0.0100
	DEWS ≥ 3	0.5%	99.9%	75.3%	77.0%	0.0100
	DEWS ≥ 7.1	0.8%	99.9%	63.0%	87.0%	0.0150
	DEWS ≥ 8.0	0.8%	99.9%	60.7%	88.3%	0.0160
	DEWS ≥ 18.2	1.4%	99.9%	49.3%	94.6%	0.0280
	DEWS ≥ 52.8	3.7%	99.9%	37.3%	98.4%	0.0710
Ueno Ryo et al. [18]	RF (medical patient, biosignal data)	4.7%	99.7%	80.3%	78.3%	0.0888
	RF (medical patient, biosignal, and laboratory data)	5.2%	99.7%	79.6%	80.9%	0.0976
	RF (surgical patient, biosignal data)	2.0%	99.8%	70.8%	81.9%	0.0389
	RF (surgical patient, biosignal, and laboratory data)	1.8%	99.8%	70.7%	79.5%	0.0351
	RF (ICU patient, biosignal data)	34.6%	90.5%	88.9%	38.4%	0.4981
	RF (ICU patient, biosignal, and laboratory data)	38.2%	85.9%	72.5%	56.0%	0.5004
	RF (ward patient, biosignal data)	2.2%	99.9%	81%	79.1%	0.0428
	RF (ward patient, biosignal, and laboratory data)	2.4%	99.9%	78.2%	81.4%	0.0466
Our previous research [23]	DT	46.80%	99.01%	28.99%	99.54%	0.3580
	RF	98.22%	98.95%	24.25%	100.00%	0.3894
	LR	5.14%	99.57%	76.33%	80.35%	0.0964
	LSTM model	38.37%	99.06%	32.66%	99.27%	0.3528
	GRU model	34.59%	99.09%	34.59%	99.09%	0.3469
	LSTM–GRU hybrid model	30.53%	99.14%	38.65%	98.77%	0.3412
Our proposed methods	DT	75.80%	99.28%	57.02%	99.69%	0.6508
	RF	96.88%	99.24%	54.39%	99.97%	0.6966
	LR	23.84%	99.81%	89.04%	95.22%	0.3761
	LSTM model	85.92%	99.83%	89.69%	99.75%	0.8777
	GRU model	84.95%	99.84%	90.35%	99.73%	0.8757
	LSTM–GRU hybrid model	79.76%	99.80%	88.16%	99.62%	0.8375

SPTTS: single-parameter track and trigger system; MEWS: the modified early warning score; DEWS: deep learning-based early warning system; DT: decision tree; RF: random forest; LR: logistic regression; LSTM: long short-term memory; GRU: gated recurrent unit; PPV: positive predictive value; NPV: negative predictive value.

Traditional EWS is useful for classifying emergency patients [6] but there is a problem with false alarms in predicting cardiac arrest. DEWS performs prediction through a deep learning algorithm based on biosignal data [6] which we will apply to hospitals. Therefore, we considered laboratory data. Ueno Ryo et al. studied the random forest algorithm that showed low PPV [18]. Our proposed methods show higher PPV and sensitivity.

We confirmed that early cardiac arrest prediction models based on deep learning have high PPV. In our previous study, we excluded patients without laboratory data and performed evaluation. We confirmed that the check of the measurement of laboratory data did not train as intended. We only used 8 h of past patient data in the previous study [23]. In this study, we excluded a patient whose laboratory data were not measured. We changed the past patient range because we improved the PPV and sensitivity. However, our proposed methods are difficult to apply to patients whose data range is under 64 h.

The dataset in our research had three limitations. First, each patient had different measurement interval times because each patient's data were measured by the medical staff who considered the patient's individual condition. However, deep learning based on the RNN model requires the same interval time. We changed the patient measurement interval time to 1 h. We replaced the missing values and set last measure values. Second, our datasets were only collected from Soonchunhyang University Cheonan Hospital. In this research, the population was therefore homogenous. Third, we did not consider patient data which were less than 64 h in the test dataset. Therefore, it is difficult to apply our method to patients with short-stay patients.

Deep learning methods based on RNN models can potentially predict cardiac arrest at an early stage using time-series data. In this research, we predict cardiac arrest within 8 h. The cardiac arrest early prediction algorithm has potential within 1 h in the future. In this research, we only used 72 h of patient data. When using real-time data from hospitals without data preprocessing, we have to solve the problem of low PPV and sensitivity.

6. Conclusions

We proposed in-hospital cardiac arrest prediction models in Soonchunhyang University Cheonan Hospital based on machine learning. We demonstrated improved performance based on input parameters (e.g., consider: maximum SBP, minimum SBP; ignore: sex, DBP, AST). We demonstrated improved performance based on the number of past patient data (64 h). We demonstrated improved performance based on hyperparameter tuning (decision tree: k set 10; random forest: k set 10; logistic regression: k set 4; LSTM model: the unit size of the layer set 96; GRU model: the unit size of the layer set 96; LSTM–GRU hybrid model: the unit size of the layer set 64). Our proposed methods only predict cardiac arrest or lack of cardiac arrest. In the future, we plan to develop a scoring system for medical staff to confirm cardiac arrest risk levels and an explainable artificial intelligence (XAI) model for cardiac arrest prediction. In the future, we aim to predict the cardiac arrest forecast time such as within 1 h, 2 h, 4 h, 8 h, 12 h, and 16 h. We will test our machine learning algorithms in Soonchunhyang University Cheonan Hospital, Soonchunhyang University Bucheon Hospital, Soonchunhyang University Seoul Hospital, and Soonchunhyang University Gumi Hospital.

Author Contributions: Conceptualization, H.L. and H.-W.G.; methodology, M.C. and H.L.; software, M.C.; validation, N.-J.C., H.-W.G. and H.L.; formal analysis, M.C.; investigation, M.C.; resources, H.-W.G. and N.-J.C.; data curation, N.-J.C.; writing—original draft preparation, M.C.; writing—review and editing, H.L. and H.-W.G.; visualization, M.C.; supervision, H.L.; project administration, H.-W.G.; funding acquisition, H.-W.G. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Bio & Medical Technology Development Program and Basic Science Research Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. NRF-2019M3E5D1A02069073 and NRF-2021R1A2C1009290).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Soonchunhyang University Cheonan Hospital (Cheonan, Korea) (IRB-No: 2020-02-016, approval date: 16 February 2020).

Informed Consent Statement: Patient consent was waived because of the retrospective design of the study.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. About Cardiac Arrest. Available online: https://www.heart.org/en/health-topics/cardiac-arrest/about-cardiac-arrest (accessed on 2 March 2021).
2. An, J.; Kweon, S.; Yoon, H. *Incidences of Sudden Cardiac Arrest in Korea, 2019*; Korea Disease Control and Prevention Agency: Seoul, Korea, 2021.
3. Kang, J.; Kim, Y.; Shin, Y.; Huh, J.; Hong, S.; Kim, W. Association Between Time to Defibrillation and Neurologic Outcome in Patients with In-Hospital Cardiac Arrest. *Am. J. Med. Sci.* **2019**, *8*, 143–148. [CrossRef] [PubMed]
4. Vähätalo, J.H.; Huikur, H.V.; Holmström, L.T.A.; Kenttä, T.V.; Haukilaht, M.A.E.; Pakanen, L.; Kaikkonen, K.S.; Tikkanen, J.; Perkiömäki, J.S.; Myerburg, R.J.; et al. Association of Silent Myocardial Infarction and Sudden Cardiac Death. *JAMA Cardiol.* **2019**, *4*, 796–802. [CrossRef] [PubMed]
5. Miyazaki, A.; Sakaguchi, H.; Ohuchi, H.; Yasuda, K.; Tsujii, N.; Matsuoka, M.; Yamamoto, T.; Yazaki, S.; Tsuda, E.; Yamada, O. The clinical characteristics of sudden cardiac arrest in asymptomatic patients with congenital heart disease. *Heart Vessel.* **2015**, *30*, 70–80. [CrossRef] [PubMed]
6. Kwon, J.; Lee, Y.; Lee, Y.; Lee, S.; Park, J. An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest. *J. Am. Heart Assoc.* **2018**, *7*, e008678. [CrossRef]
7. Yoo, H.; Park, R.; Chung, K. IoT-Based Health Big-Data Process Technologies: A Survey. *KSII Trans. Internet Inf. Syst.* **2021**, *15*, 974–992.
8. Mohemmed, S.M.; Parveen, R.M. Cloud-based Healthcare data management Framework. *KSII Trans. Internet Inf. Syst.* **2020**, *14*, 1014–1025.
9. Wu, P.; Li, H.; Zeng, N.; Li, F. FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. *Image Vis. Comput.* **2022**, *117*, 104341. [CrossRef]
10. Zeng, N.; Li, H.; Peng, Y. A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease. *Neural Comput. Appl.* **2021**, 1–12. [CrossRef]
11. Zeng, N.; Wang, Z.; Zhang, H.; Kim, K.-E.; Li, Y.; Liu, X. An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *IEEE Trans. Nanotechnol.* **2019**, *18*, 819–829. [CrossRef]
12. Florence, D.; Wulfran, B.; Alain, C. Cardiac arrest: Prediction models in the early phase of hospitalization. *Curr. Opin. Crit. Care* **2019**, *25*, 204–210.
13. Somanchi, S.; Adhikari, S.; Lin, A.; Eneva, E.; Ghani, R. Early Prediction of Cardiac Arrest (Code Blue) using Electronic Medical Records. In Proceedings of the 21 th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 2119–2126.
14. Ong, M.E.H.; Ng, C.H.L.; Goh, K.; Liu, N.; Koh, Z.X.; Shahidah, N.; Zhang, T.T.; Chong, S.F.; Lin, Z. Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score. *Crit. Care* **2012**, *16*, R108. [CrossRef]
15. Churpek, M.M.; Yuen, T.C.; Park, S.; Meltzer, D.O.; Hall, J.B.; Edelson, D.P. Derivation of a cardiac arrest prediction model using ward vital signs. *Crit. Care Med.* **2013**, *40*, 2102–2108. [CrossRef]
16. Liu, N.; Lin, Z.; Cao, J.; Koh, Z.; Zhang, T.; Huang, G.; Ser, W.; Ong, M.E.H. An Intelligent Scoring System and Its Application to Cardiac Arrest Prediction. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 1324–1331. [CrossRef]
17. Murukesan, L.; Murugappan, M.; Iqbal, M.; Saravanan, K. Machine Learning Approach for Sudden Cardiac Arrest Prediction Based on Optimal Heart Rate Variability Features. *J. Med. Imaging Health Inform.* **2014**, *4*, 521–532. [CrossRef]
18. Ueno, R.; Xu, L.; Uegami, W.; Matsui, H.; Okui, J.; Hayashi, H.; Miyajima, T.; Hayashi, Y.; Pilcher, D.; Jones, D. Value of laboratory results in addition to vital signs in a machine learning algorithm to predict in-hospital cardiac arrest: A single-center retrospective cohort study. *PLoS ONE* **2020**, *15*, e0235835. [CrossRef]
19. Sbrollini, A.; Jongh, M.C.D.; Haar, C.C.T.; Treskes, R.W.; Man, S.; Burattini, L.; Swenne, C.A. Serial electrocardiography to detect newly emerging or aggravating cardiac pathology: A deep-learning approach. *BioMed. Eng. OnLine* **2019**, *18*, 15. [CrossRef]
20. Ibrahim, L.; Mesinovic, M.; Yang, K.W.; Eid, M.A. Explainable Prediction of Acute Myocardial Infarction Using Machine Learning and Shapley Values. *IEEE Access* **2020**, *8*, 210410–210417. [CrossRef]
21. ElSaadany, Y.; Majumder, A.J.A.; Ucci, D.R. A Wireless Early Prediction System of Cardiac Arrest through IoT. In Proceedings of the 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, Italy, 4–8 July 2017; pp. 690–695.

22. Churpek, M.M.; Yuen, T.C.; Huber, M.T.; Park, S.Y.; Hall, J.B.; Edelson, D.P. Predicting Cardiac Arrest on the Wards: A Nested Case-Control Study. *Chest* **2015**, *141*, 1170–1176. [CrossRef]
23. Chae, M.; Han, S.; Gil, H.; Cho, N.; Lee, H. Prediction of In-Hospital Cardiac Arrest Using Shallow and Deep Learning. *Diagnostics* **2021**, *17*, 1225. [CrossRef]
24. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
25. Chollet, F. Keras GitHub Repository. 2015. Available online: https://github.com/fchollet/keras (accessed on 2 March 2020).
26. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
27. Coppersmith, D.; Hong, S.; Hosking, J.R.M. Partitioning Nominal Attributes in Decision Trees. *Data Min. Knowl. Discov.* **1999**, *3*, 197–217. [CrossRef]
28. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; pp. 278–282.
29. Cramer, J.S. The Origins of Logistic Regression. *Tinbergen Institute Discussion Papers 02-119/4*. 2002. Available online: https://ideas.repec.org/p/tin/wpaper/20020119.html (accessed on 2 March 2021).
30. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [CrossRef]
31. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
32. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
33. Cho, K.; Merrienboer, B.V.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceeding of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1724–1734.
34. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]