

RESEARCH

Open Access



# Revisiting the potential value of vital signs in the real-time prediction of mortality risk in intensive care unit patients

Pan Pan<sup>1†</sup>, Yue Wang<sup>2†</sup>, Chang Liu<sup>1</sup>, Yanhui Tu<sup>2,3</sup>, Haibo Cheng<sup>2,3</sup>, Qingyun Yang<sup>1</sup>, Fei Xie<sup>1</sup>, Yuan Li<sup>4</sup>, Lixin Xie<sup>1\*†</sup> and Yuhong Liu<sup>1\*†</sup>

<sup>†</sup>Pan Pan and Yue Wang are first authors and contributed equally to this article.

<sup>†</sup>Yuhong Liu and Lixin Xie are both corresponding authors and contributed equally to this article.

\*Correspondence:  
xielx301@126.com;  
doctorliu1974@163.com

<sup>1</sup> College of Pulmonary & Critical Care Medicine, 8th Medical Center, Chinese PLA General Hospital, Beijing, China

<sup>2</sup> Shandong Future Network Research Institute, Jinan, China

<sup>3</sup> Jiangsu Future Network Group Co, Nanjing, China

<sup>4</sup> Department of Computer, Chinese PLA General Hospital, Beijing, China

## Abstract

**Background:** Predicting patient mortality risk facilitates early intervention in intensive care unit (ICU) patients at greater risk of disease progression. This study applies machine learning methods to multidimensional clinical data to dynamically predict mortality risk in ICU patients.

**Methods:** A total of 33,798 patients in the MIMIC-III database were collected. An integrated model NIMRF (Network Integrating Memory Module and Random Forest) based on multidimensional variables such as vital sign variables and laboratory variables was developed to predict the risk of death for ICU patients in four non overlapping time windows of 0–1 h, 1–3 h, 3–6 h, and 6–12 h. Mortality risk in four nonoverlapping time windows of 12 h was externally validated on data from 889 patients in the respiratory critical care unit of the Chinese PLA General Hospital and compared with LSTM, random forest and time-dependent cox regression model (survival analysis) methods. We also interpret the developed model to obtain important factors for predicting mortality risk across time windows. The code can be found in <https://github.com/wyuexiao/NIMRF>.

**Results:** The NIMRF model developed in this study could predict the risk of death in four nonoverlapping time windows (0–1 h, 1–3 h, 3–6 h, 6–12 h) after any time point in ICU patients, and in internal data validation, it is suggested that the model is more accurate than LSTM, random forest prediction and time-dependent cox regression model (area under receiver operating characteristic curve, or AUC, 0–1 h: 0.8015 [95% CI 0.7725–0.8304] vs. 0.7144 [95% CI 0.6824–0.7464] vs. 0.7606 [95% CI 0.7300–0.7913] vs. 0.3867 [95% CI 0.3573–0.4161]; 1–3 h: 0.7100 [95% CI 0.6777–0.7423] vs. 0.6389 [95% CI 0.6055–0.6723] vs. 0.6992 [95% CI 0.6667–0.7318] vs. 0.3854 [95% CI 0.3559–0.4150]; 3–6 h: 0.6760 [95% CI 0.6425–0.7097] vs. 0.5964 [95% CI 0.5622–0.6306] vs. 0.6760 [95% CI 0.6427–0.7099] vs. 0.3967 [95% CI 0.3662–0.4271]; 6–12 h: 0.6380 [0.6031–0.6729] vs. 0.6032 [0.5705–0.6406] vs. 0.6055 [0.5682–0.6383] vs. 0.4023 [95% CI 0.3709–0.4337]). External validation was performed on the data of patients in the respiratory critical care unit of the Chinese PLA General Hospital. Compared with LSTM, random forest and time-dependent cox regression model, the NIMRF model was still the best, with an AUC of 0.9366 [95% CI 0.9157–0.9575 for predicting death risk in 0–1 h]. The corresponding AUCs of LSTM, random forest and time-dependent cox regression

model were 0.9263 [95% CI 0.9039–0.9486], 0.7437 [95% CI 0.7083–0.7791] and 0.2447 [95% CI 0.2202–0.2692], respectively. Interpretation of the model revealed that vital signs (systolic blood pressure, heart rate, diastolic blood pressure, respiratory rate, and body temperature) were highly correlated with events of death.

**Conclusion:** Using the NIMRF model can integrate ICU multidimensional variable data, especially vital sign variable data, to accurately predict the death events of ICU patients. These predictions can assist clinicians in choosing more timely and precise treatment methods and interventions and, more importantly, can reduce invasive procedures and save medical costs.

**Keywords:** Real-time prediction, Risk of death, Machine learning, Predictive models, ICU

## Introduction

Prediction of mortality risk in ICU patients is an important topic in ICU clinical practice. Effectively predicting the risk of death of patients can assist clinicians in taking reasonable treatment and intervention measures earlier and improve the prognosis of patients. Modeling the risk of death in patients was performed more than half a century ago: the Apgar score for assessing neonatal risk [1] was first published in 1952, followed by the introduction of acute physiology, age, and chronic health assessment (APACHE scoring system) in 1981 by the Kanus et al. [2] and the widely used Simplified Acute Physiology Score (SAPS) [3] was released in 1984. However, these scoring systems and prognostic models are defined by physicians based on their own experience, selected patients, and statistical analysis. However, the treatment methods and levels of treatment vary in different countries and regions, resulting in the above scoring system being less indicative of prognosis [4–10]. The Simplified Acute Physiology Score-III (SAPS-III) was developed on a global scale to avoid regional differences, providing a custom formula to correlate risk-adjusted expected mortality with ward location, but risk-adjusted mortality predictions include overestimation of expected mortality [10, 11] or underestimation of observed mortality [12]. In addition, the occurrence and development of disease in critically ill patients is an extremely rapid pathophysiological process rather than a static state. It is not possible to simply observe and measure at a time point to explain the current situation of the patient or the response to treatment. Faced with changing baseline characteristics, therefore, designers of scoring systems need to review models at regular intervals or define increasingly complex scores in pursuit of high accuracy, such as the Acute Physiology and Chronic Health Assessment (APACHE) IV [11]. The clinical variables required for scoring are almost twice those of APACHE II. This is undoubtedly unrealistic in actual clinical implementation, so we urgently need a real-time tool that can objectively assess the risk of death of patients.

Machine learning enables to learn from data and make predictions, and the learning performance increases with the training data [13]. In terms of predicting the risk of death for ICU patients, machine learning technology has better predictive performance than traditional scoring methods such as APACHE and SAPS [14]. At present, some studies have been used to predict the prognosis of ICU patients through machine learning. For example, Gong et al. used logistic regression to predict patient in-hospital mortality and prolonged hospital stay based on electronic health records (EHRs) and finally demonstrated that for both outcomes, EHR-specific events were mapped to a set

of shared clinical concepts of features, indicating that it yielded better results than using EHR-specific events alone [15]. Recent studies have shown that novel neural architectures (including LSTM-based neural architectures) perform well in predicting mortality in hospitalized patients with an AUC value of 0.93 [16].

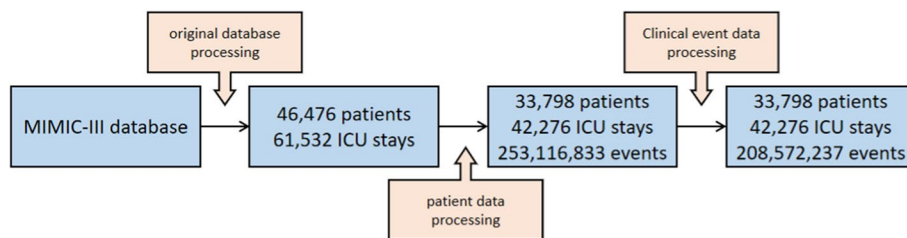
MIMIC is a large, freely usable clinical intensive care database published by the Computational Physiology Laboratory of the Massachusetts Institute of Technology. It records clinical data of patients admitted to the Beth Israel Deaconess Medical Center, including detailed information on patient demographics, laboratory tests, medication use, vital signs, surgical procedures, disease diagnosis, drug management, and survival status. It now includes versions II, III, and IV. Due to the large sample size, comprehensive information, and long patient tracking time of MIMIC, many ICU mortality risk prediction studies are based on this database. Ghassemi et al. [17] used a latent variable model to decompose the free text hospital records of MIMIC-II into meaningful features, predicting in-hospital, 30 days after discharge, and 1 year after discharge mortality rates. A retrospective model combining potential thematic and structural features predicted in-hospital, 30 days after discharge, and 1 year mortality rates with AUCs of 0.96, 0.82, and 0.81, respectively. Zhang et al. [18] extracted non-invasive variable data from MIMIC-III that can be obtained through monitors and manual measurements, and trained four machine learning models to predict 28-day mortality. Among them, the optimal model based on the LightGBM algorithm showed good performance, with an accuracy of 0.797 and an AUC of 0.879. However, most machine learning models for predicting the death risk of ICU patients predict 28-day mortality or in-hospital mortality, and it is difficult to achieve real-time dynamic prediction. Many models use a small number of features for prediction [19]. In fact, using fewer features in prediction is also in line with the goal of making faster predictions for ICU patients, resulting in a better prediction effect. In addition, most of the published machine learning studies on ICU patients are based on the MIMIC database [20], and there are few studies on Chinese ICU data or are based on the data of inpatients in the intensive care unit of a domestic hospital, lacking external validation [21]. This study aims to build a model based on MIMIC data and multidimensional data such as vital sign variables and laboratory variables to dynamically predict the mortality risk of ICU patients after any time point in real-time. The model is interpreted to better understand the contribution of predictive models and features.

## Method

### Study population and dataset

Our study was based on the MIMIC-III critical care database. First, we define some terms: in MIMIC-III, each patient has one or more hospital admissions; during an admission, a patient may have one or more ICU stays (ICU stays), which we call fragments (episodes); and a clinical event refers to a measurement, observation, or treatment of a patient. A sample is a single record processed by the model, and a sample in this study is an event that occurred within a window of observational data before the time of interest.

The data preparation process is shown in Fig. 1, including three parts: original database sorting, patient data processing, and clinical event data processing. (1) Organizing



**Fig. 1** Data preparation flow chart

the original database. Relevant data were extracted from the original MIMIC-III intensive care database and organized by patient, including more than 60,000 ICU admissions for more than 40,000 intensive care patients. (2) Patient data processing. Exclusion criteria were used for admissions and ICU admissions: admitted patients transferred to different general wards or ICU wards and patients with multiple ICU admissions were excluded to avoid ambiguity in the analysis results related to admissions rather than ICU admissions, taking into account differences in physiology between adults and children, and excluding all ICU admissions younger than 18 years. The resulting root cohort contained 33,798 patients, a total of 42,276 ICU admissions, and more than 250 million clinical events. (3) Data processing of clinical events. Remove more than 44.5 million events that cannot be reliably matched to ICU admissions in the root cohort: first, remove events without an admission ID (HADM\_ID), events with an admission ID but not in stays.csv, where stays.csv is a table that associates ICU admission attributes (such as length of stay and mortality) with admission IDs; second, for events with missing ICU admission IDs (ICUSTAY\_ID), reliable recovery of ICU admission IDs by admission ID; and finally, deletion of ICU admission IDs Events not listed in stays.csv.

After the data preparation is completed, the validation set and the test set are divided according to the proportion of 15%. In addition, the same process was used to process the electronic case data from the PLA General Hospital for the validation work in the study.

**Variable selection**

We have reviewed the physiological variables used in existing studies on patient mortality risk (Table 1), and summarized clinically significant variables suggested by multiple doctors from different hospitals to generate candidate predictive variables. Based on the training set data, the pairwise Pearson correlation coefficients of the candidate predictors are calculated to analyze the correlation between the variables. A predefined predictor is generated by removing one of the two predictors with a correlation coefficient greater than 0.6 to avoid redundancy in the model variables.

**Sample generation**

The risk of death within a time window is usually defined as a dichotomous classification based on data from a limited period of time after ICU admission, with data labels indicating whether a patient died within a certain time window after the moment of interest. The general model uses the data of the first period of time after admission to the ICU. We use the data of a certain observation time window before the time of interest to

**Table 1** Variables used in studies related to mortality risk of ICU patients (6 categories)

Category	Variable	References	Static/dynamic
Blood gas	(A-a)DO <sub>2</sub>	[2, 23]	Dynamic
Blood gas	FIO <sub>2</sub>	[22]	Dynamic
Blood gas	HCO <sub>3</sub>	[2, 23]	Dynamic
Blood gas	Oxygenation index (P/F)	[3, 19, 23]	Dynamic
Blood gas	PaO <sub>2</sub>	[2, 23]	Dynamic
Blood gas	Potassium	[2, 3, 22, 23]	Dynamic
Blood gas	Sodium	[2, 3, 22, 23]	Dynamic
Blood gas	SpO <sub>2</sub>	[18, 22]	Dynamic
Blood gas	PH (arterial blood)	[2, 23]	Dynamic
Demographic	Age	[2, 3, 15, 18–22]	Static
Demographic	Appearance	[1]	Static
Demographic	Comorbid	[20, 23]	Static
Demographic	ICU type	[3, 15, 21]	Static
Demographic	Sex	[15, 20–22]	Static
Hemodynamic	BSA	[19]	Dynamic
Hemodynamic	Cardiac index	[23]	Dynamic
I/O	Input (24 h)	[19]	Dynamic
I/O	Output (24 h)	[19]	Dynamic
I/O	Urine volume	[3, 22]	Dynamic
Lab	Actual bicarbonate	[3]	Dynamic
Lab	Albumin	[19]	Dynamic
Lab	Blood lactic acid	[19]	Dynamic
Lab	Blood urea nitrogen	[3]	Dynamic
Lab	Chloride	[22]	Dynamic
Lab	Creatinine	[2, 17, 22, 23]	Dynamic
Lab	Glucose	[22, 23]	Dynamic
Lab	Hemoglobin	[19, 23]	Dynamic
Lab	Platelet count	[19, 22, 23]	Dynamic
Lab	Total bilirubin	[3, 19, 23]	Dynamic
Lab	White blood cell count	[2, 3, 17, 23]	Dynamic
Lab	White blood cell percentage	[2]	Dynamic
Vital	Apgar activity	[1]	Dynamic
Vital	Apgar grimace	[1]	Dynamic
Vital	Apgar respiration	[1]	Dynamic
Vital	BMI	[22]	Dynamic
Vital	CHS	[2, 3]	Dynamic
Vital	DBP	[3, 15, 17, 18, 21–23]	Dynamic
Vital	GCS	[2, 3, 22, 23]	Dynamic
Vital	Height	[15, 19, 21]	Dynamic
Vital	HR	[2, 3, 17–19, 22, 23]	Dynamic
Vital	MBP	[2, 15, 18, 19, 21–23]	Dynamic
Vital	NEWS	[17]	Dynamic
Vital	Pulse	[1]	Dynamic
Vital	RASS	[23]	Dynamic
Vital	RR	[2, 17, 18, 22, 23]	Dynamic
Vital	SAPS-II	[20]	Dynamic
Vital	SBP	[15, 18, 21–23]	Dynamic
Vital	Temperature	[2, 3, 17–19, 22, 23]	Dynamic
Vital	Weight	[15, 18, 19, 21]	Dynamic

obtain more real-time dynamic information and choose the observation time window to be 48 h to ensure that the detection is more likely to reflect the patient's characteristics of disease changes.

To prepare training samples for mortality risk prediction within a certain time window, first, compile the time series of events for each segment processed by the process in Fig. 1 and only retain the predefined model variable data. Second, slice the time series to generate samples and corresponding labels, record the start time of a patient sample as  $t_0$  and the end time as  $t_1$ . Then, the patient sample is the time series data in the time period of  $(t_0, t_1)$ . For the patient sample and a certain time window  $(h_0, h_1)$  (for example, to predict the risk of death within 1 to 3 h from a patient's time of interest, the time window is  $(1, 3)$ ). If the patient dies within the  $(t_1 + h_0, t_1 + h_1)$  time period, then this sample is a positive sample for predicting the risk of death within that time window, labeled as "death within this time window". If the patient survives or dies outside the  $(t_1 + h_0, t_1 + h_1)$  time period, then this sample is a negative sample for predicting the risk of death within that time window, labeled as "death outside this time window". Then, the sample is a negative sample for the prediction of death risk in this time window, and the label is "death outside this time window". The specific sample generation rules are as follows: (1) based on the time of death of the dead patient in the ICU, a certain 48-h data is intercepted forward as a positive sample, where the 48-h data start time = death time - random time period - 48 h, data end time = death time - random time period, the random time period corresponds to the time window, and is a random number in the time window interval, to ensure that the positive samples that died within the time window are generated; (2) For patients who did not die in the ICU, extract the data within 48 h after admission to the ICU is used as a negative sample; (3) For the segments with time series data less than 48 h extracted based on the above two rules, we fill in the data to 48 h. For details, see the description in "Variable data preprocessing" section.

Real-time and accurate monitoring and condition judgment of critically ill patients are of great significance. According to clinical guidelines, patient treatment usually needs to complete the corresponding goals within 1 h, 3 h, and 6 h. Therefore, this study selected 0–1 h, 1–3 h, 3–6 h, and 6–12 h as the preliminary research time window. Unless otherwise specified, "1 h", "3 h", "6 h" and "12 h" mentioned later represent 0–1 h, 1–3 h, 3–6 h and 6–12 h, respectively.

To better predict the death risk of patients in different time windows, we hope to reduce the misidentification between death samples in different time windows. Therefore, for negative samples corresponding to a certain time window, in addition to the aforementioned negative samples generated by patients who did not die in the ICU, samples that died in other time windows were generated by deceased patients in the ICU, and the positive samples corresponding to this time window were the previously generated samples that died in this time window.

### Variable data preprocessing

The different data units, missing data, and non-numerical variables of clinical data make difficulties in establishing machine learning models. Data preprocessing is a necessary for making raw data suitable for predictive model development. It includes five steps:

data assembly, missing value filling, One-Hot encoding, data normalization, and data mapping.

**Data assembly:** The finalized predictor data were compiled from multiple MIMIC-III variables (Additional file 1: Appendix S1). For example, there are 8 variables in the chart events table (ITEM\_ID are 3655, 677, 676, 223762, 3654, 678, 223761, 679) corresponding to body temperature, these variables have different units (Celsius °C, Fahrenheit °F), we put this All values of the 8 variables are converted to data in degrees Celsius. Similar summarization and unit conversion preprocessing were also performed on the other 42 variables to maintain the consistency of the data distribution of the same variables.

**Missing value filling:** The data sample in this study needs to have at least one piece of data per hour to improve data quality. The statistics of the proportion of missing data for each variable are shown in Additional file 1: Appendix S2. For time series data samples of less than 48 h, the first and last data are used to copy and complete them at the beginning and the end, respectively; for missing variable data, the previous data at the nearest time point are used for filling. If all variable data were missing, a normal value was selected from the normal range of the variable to fill in.

**Data mapping:** Based on the extracted 48-h time series sample data, divide it into 48 time periods in units of 1 h, and sequentially take the latest collected predictive variable data from each time period as the corresponding data for that time period. The final generated multi-dimensional variable data slice sample is a  $48 \times 43$  dimensional matrix. The number of rows 48 represents the total amount of data in the 48 h slice sample, and the number of columns 43 corresponds to the number of predicted variables.

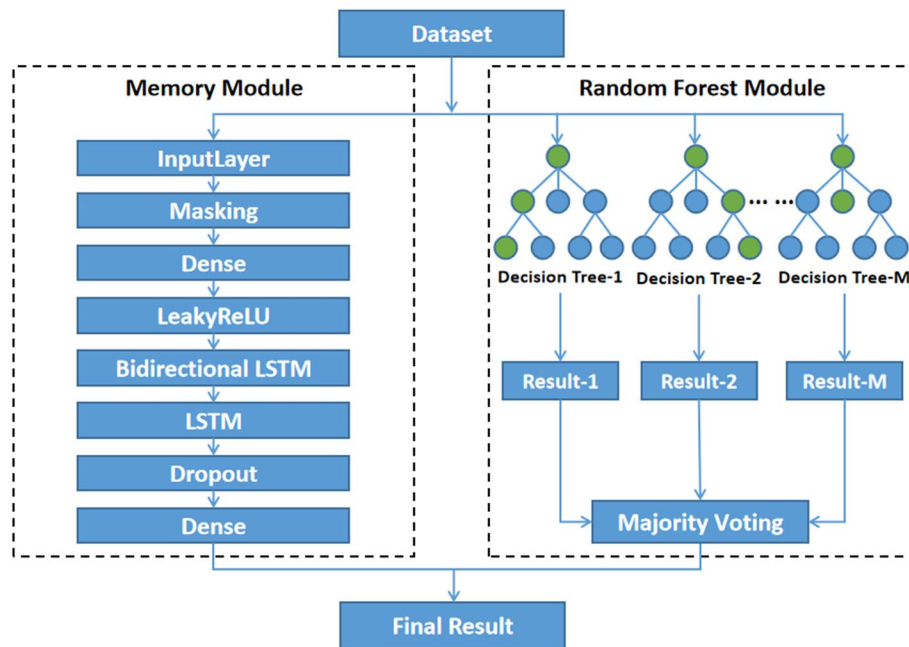
**One-Hot encoding:** Using One-Hot encoding to encode type variables as state values for further data analysis [22]. In this study, the type variable is stool color (including 10 categories: Black, BrightRedBlood, Brown, Clay, Clear, Golden, Green, Maroon, Melena, Others).

**Data normalization:** Divide the data of each variable by its maximum absolute value, and scale it to a range (0, 1) to reduce the adverse impact of dimensional differences on the algorithm.

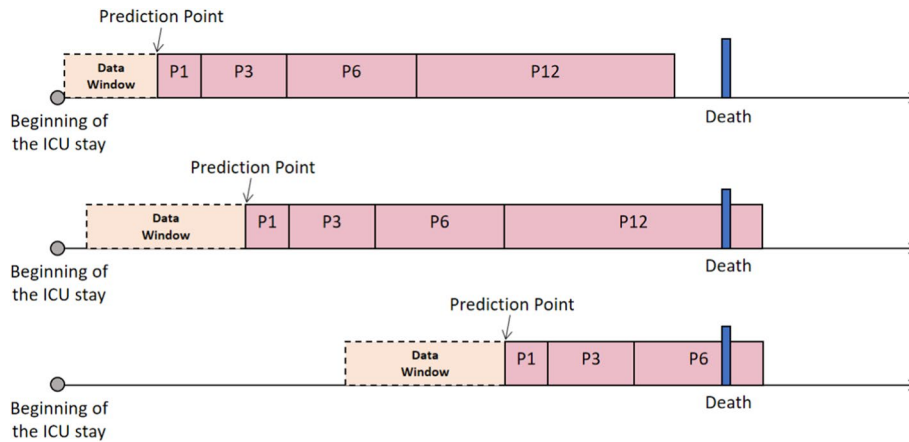
### **Model development**

The samples in this study are model variable data from a certain 48 h time period. In practical clinical practice, some variable data has a huge amount and are collected with small time intervals (such as respiration, heart rate, etc.), thus can be considered as time series data within 48 h. However, other part of the variable data has a long-time interval for collection, with sparse data within 48 h, making it difficult to form a time series. We designed memory modules and random forest modules respectively for the characteristics of time series data and non-time series data. The memory module fully learns the temporal information of the data, and the random forest module simulates clinical diagnosis to improve algorithm accuracy and increase model interpretability. The integration of two modules has constructed a NIMRF model (Network Integrating Memory Module and Random Forest) that can fully learn time series and non-time series information, as shown in Fig. 2. The above two sub modules respectively use LSTM [23] and random forest [24] as the backbone networks, and make appropriate adjustments and improvements based on the characteristics of the mortality risk prediction task in this





**Fig. 2** NIMRF network structure diagram



**Fig. 3** Schematic diagram of study design and model development. For the time of interest (Prediction Point), the variable data in the observation data window (Data Window) before the time are collected to predict the death risk at this time. P1: 1-h mortality risk prediction; P3: 3-h mortality risk prediction; P6: 6-h mortality risk prediction; P12: 12-h mortality risk prediction

study (such as complex sample data and easy overfitting) on the basis of the backbone network. The internal and external validation results demonstrate that NIMRF has relatively accurate predictive performance, and the details of these architectures are explained in Additional file 1: Appendix S3. The study design and model development are shown in Fig. 3. For each moment of interest, we developed four NIMRF models to predict a patient’s risk of death 1, 3, 6, and 12 h after that moment. In addition, we also used LSTM and random forest algorithm models to predict the mortality risk in the above four-time windows.



### Model comparison

Firstly, three machine learning algorithms are briefly introduced: (1) LSTM [23] (Long Short Term Memory Network) is a special type of RNN (Recurrent Neural Network, a type of neural network used to process sequence data), which uses several structures called “thresholds” to manage information transmission, selectively allowing information to pass through. Therefore, it can perform better in longer sequences compared to ordinary RNNs; (2) Random Forest [24] is an ensemble learning algorithm that uses decision trees as the basic unit and combines the prediction results of multiple decision trees to obtain the final prediction result. (3) The Cox proportional hazards model [25] is a commonly used survival analysis method that uses final outcome and survival time as dependent variables to calculate survival probabilities at different times.

Based on the model variables in this study, in addition to NIMRF, we developed four models using LSTM and random forest algorithm for predicting the risk of death within the four-time windows mentioned above. Considering that our study is a time-varying event, we also attempted to develop a Cox regression model based on survival analysis. Through Schoenfeld residual validation, we found that some variables in this study did not meet the proportional risk hypothesis required by Cox regression (proportional risk hypothesis test is in Additional file 1: Appendix S4). Therefore, we introduced the time-dependent variable  $x \cdot \log(t + 20)$  into the Cox model to address this issue and also developed a Cox regression model with time-dependent covariates for prediction. As for NIMRF, we also trained models for prediction based on the other two variable combination schemes: (1) based on the variable combination age, sex, Vent, BMI, urine output, GCS, FIO<sub>2</sub>, HR, RR, T, SPO<sub>2</sub>, SBP, DBP, MBP, chlorine, creatinine, glucose, potassium, sodium, platelet count, pH, as described in Reference 18; (2) all vital sign variables selected in our study are SBP, DBP, HR, RR, and Temperature.

The output of the model developed above is the probability of patient death predicted by the model in the corresponding time window. Based on this, the performance of the model is statistically analyzed and compared.

### Statistical analysis

We calculated the AUC (area under the receiver operating characteristic curve) based on the aforementioned divided MIMIC test set data and the electronic case data of the Chinese People’s Liberation Army General Hospital from 2007 to 2016 to evaluate the performance of the prediction model and calculated the 95% confidence of the AUC by the bootstrapping method. In addition, the sensitivity and specificity of the model under different thresholds were calculated.

### Model interpretation

For the memory module of the model, we use a fully connected layer after the input layer to represent the variable distribution of the input and then use the Leaky-ReLU activation layer to increase the nonlinearity of the model. At this time, for each variable, the corresponding fully connected layer is calculated. The weight parameter average reflects the contribution of each variable in the model prediction. For the random forest module of the model, the Scikit-learn library is used to calculate

**Table 2** Statistics of MIMIC-III sample characteristics

	<b>Training set n = 23,556</b>	<b>Validation set n = 5070</b>	<b>Test set n = 5172</b>
Age, mean (SD)	64 (18)	64 (18)	64 (18)
Gender			
Female, n (%)	10,278 (44)	2230 (44)	2251 (44)
Male, n (%)	13,278 (56)	2840 (56)	2921 (56)
Race			
Asian, n (%)	578 (2.45)	111 (2.19)	132 (2.55)
Black, n (%)	1857 (7.88)	403 (7.95)	381 (7.37)
Hispanic/Latino, n (%)	782 (3.32)	174 (3.43)	165 (3.19)
Other/unknown, n (%)	3573 (15.17)	765 (15.09)	790 (15.27)
White, n (%)	16,766 (71.18)	3617 (71.34)	3704 (71.62)
Deaths in ICU, n (rate per 1000 admissions)	2006 (85)	407 (80)	463 (90)

**Table 3** Hospital patient characteristics statistics

	<b>Universal set n = 889</b>	<b>Subset n = 605</b>
Age, mean (SD)	66.80 (17.36)	65.30 (18.32)
Gender		
Female, n (%)	264 (29.70)	264 (43.64)
Male, n (%)	625 (70.30)	341 (56.36)
Race		
Asian, n (%)	889 (100)	605 (100)
Deaths in ICU, n (rate per 1000 admissions)	116 (130)	52 (86)

and visualize the contribution of each variable in the random forest model. Using the SHAP method to interpret the model output, studying the importance and positive negative relationships of each variable in predicting mortality risk in the NIMRF model through global interpretation, and obtaining the role of each variable in individual prediction through local interpretation.

## Result

### Study population characteristics

Based on the MIMIC-III critical care database, 33,798 patients were included in the study, including 23,556 in the training set, 5070 in the validation set, and 5172 in the test set. We also conducted external verification based on the data of patients in the respiratory intensive care unit of the Chinese People's Liberation Army General Hospital (hereinafter referred to as the hospital dataset), including the electronic medical record data of a total of 889 patients from 2007 to 2016. The corresponding sample size statistics for the four-time windows are shown in Table 2. In addition, according to the proportion of positive and negative samples in the MIMIC test set, we randomly selected samples from the full set of samples generated by the hospital data to form a hospital test subset (hereinafter referred to as the hospital data subset) so

that it has a similar composition ratio to the MIMIC test set. It is convenient to compare the test results of MIMIC and hospital data. The MIMIC-III and hospital patient characteristics used in the study are shown in Tables 2 and 3.

### Model variables

The candidate predictors of the model include a total of 76 kinds of vital sign variables and laboratory variables. For the two predictors whose Pearson correlation coefficient is greater than 0.6, one of them is deleted. In this process, there are two groups of variables with strong correlation but high clinical value, and we kept these variables in the model variables. These two groups of variables are hemoglobin and red blood cell count (Pearson correlation coefficient is 0.8761), low-density cholesterol, and total calcium (Pearson correlation coefficient is 0.8795). Finally, 43 variables were retained as model variables, as shown in Table 4, and the correlation between variables is shown in Fig. 4.

### The performance comparison

In this study, we developed the NIMRF model to predict the death risk of ICU patients at 1 h, 3 h, 6 h, and 12 h after a certain time. Based on the 43 model variables in our study, we tried random forest, LSTM model and time-dependent cox regression model to predict the death risk of the corresponding time window. In addition, a prediction model was trained using NIMRF for the combination of variables in Reference 18, vital sign variables in our study, and laboratory variables in our study. The prediction performance of the seven methods is compared based on the MIMIC test set, the hospital dataset, the hospital data subset, and the full test data composed of the aforementioned MIMIC test set and hospital dataset. Among the seven methods, the NIMRF model trained on 43 predictive variables has the best predictive performance: the AUC for predicting the risk of death within 1 h on the hospital dataset is 0.9366 [95% CI 0.9157–0.9575], and the AUC for random forest, LSTM prediction is 0.7437 [95% CI respectively], 0.7083–0.7791], 0.9263 [95% CI 0.9039–0.9486] and 0.2447 [95% CI 0.2202–0.2692]. Sensitivity analysis of the model showed that the sensitivity of NIMRF was higher than that of the other six methods under the same specificity. For the full test data, when the specificity was 90%, the NIMRF model had a sensitivity of 67% for prediction of mortality risk within 1 h, compared with 62% for the LSTM, 37% for the random forest and 7% for cox regression model (Table 5). The sensitivity of the NIMRF model trained based on the variables in Reference 18 and the vital signs variables in our study were 63% and 65%, respectively. Based on four test sets, the AUC values of the six methods are detailed in Additional file 1: Appendix S5.

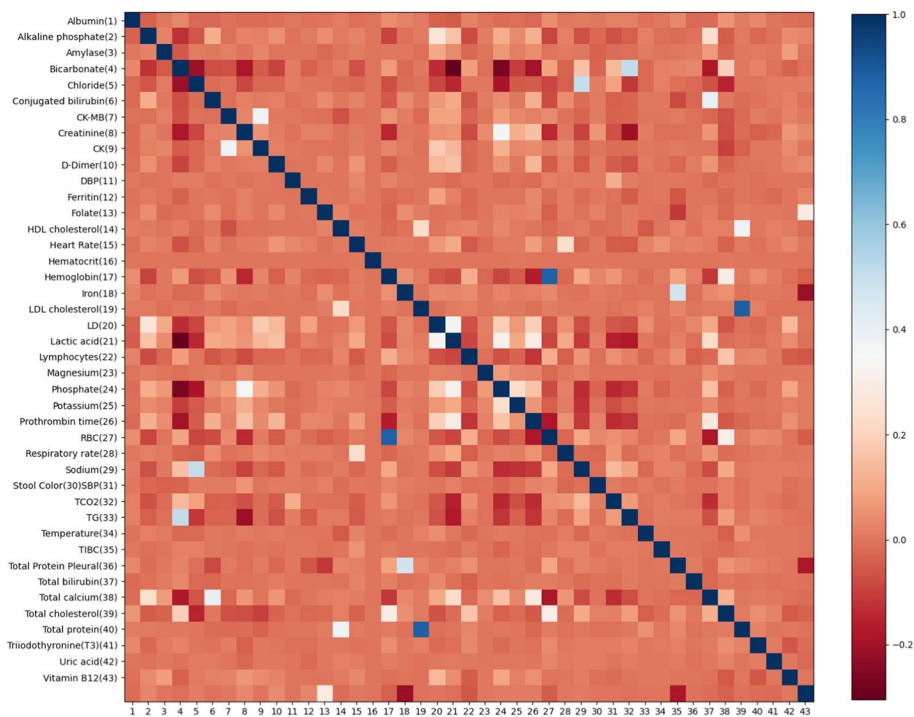
### Model validation

The NIMRF model developed in this study is validated based on the MIMIC test set, the hospital dataset, a subset of the hospital data, and the full test data. For the four-time windows of 1 h, 3 h, 6 h, and 12 h, the corresponding 4 models (NIMRF-H1, NIMRF-H3, NIMRF-H6, NIMRF-H12) on the test data are shown in Table 6. As expected, NIMRF-H1 has the best performance with an AUC of 0.9366 [95% CI 0.9157–0.9575] on the hospital dataset. The AUCs of NIMRF-H3, NIMRF-H6, NIMRF-H12 were 0.8548 [95% CI 0.8198–0.8898], 0.6780 [95% CI 0.6352–0.7207], and 0.6453 [95% CI 0.6110–0.6796],

**Table 4** Model variables

Variable	MIMIC-III table	Supplementary value	Modeled as	References
Albumin	Chartevents, labevents	4	Continuous	[19]
Alkaline phosphate	Chartevents, labevents	90	Continuous	–
Amylase	Chartevents, labevents	85	Continuous	–
Bicarbonate	Chartevents, labevents	25	Continuous	[3]
Chloride	Chartevents, labevents	101	Continuous	[2, 3, 22, 23]
Conjugated bilirubin	Chartevents, labevents	0.3	Continuous	–
Creatine kinase MB isoenzyme	Labevents	3	Continuous	–
Creatinine	Chartevents	0.9	Continuous	[2, 17, 22, 23]
Creatinine kinase(CK)	Labevents	108	Continuous	–
D-Dimer	Chartevents, labevents	300	Continuous	–
Diastolic blood pressure	Chartevents	59	Continuous	[3, 15–18, 21–23]
Ferritin	Labevents	86	Continuous	–
Folate	Labevents	9	Continuous	–
HDL cholesterol	Labevents	54	Continuous	–
Heart rate	Chartevents	86	Continuous	[2, 3, 17–19, 22, 23]
Hematocrit	Chartevents, labevents	43	Continuous	–
Hemoglobin	Chartevents, labevents	13	Continuous	[19, 23]
Iron	labevents	90	Continuous	–
LDL cholesterol	Labevents	100	Continuous	–
Lactate dehydrogenase	Labevents	180	Continuous	–
Lactic acid	Chartevents	1.1	Continuous	–
Lymphocytes	Labevents	30	Continuous	–
Magnesium	Chartevents, labevents	2	Continuous	–
Phosphate	Labevents	3.7	Continuous	–
Potassium	Chartevents, labevents	4	Continuous	[2, 3, 22, 23]
Prothrombin time	Chartevents, labevents	12.1	Continuous	–
Red blood cell count	Labevents	4.5	Continuous	[2, 3, 17, 23]
Respiratory rate	Chartevents	19	Continuous	[2, 17, 18, 22, 23]
Sodium	Chartevents, labevents	141	Continuous	[2, 3, 22, 23]
Stool color	chartevents	Golden	Categorical	–
Systolic blood pressure	Chartevents	118	Continuous	[3, 15, 17, 18, 21–23]
TCO <sub>2</sub>	Chartevents	25	Continuous	–
TG	Chartevents	125	Continuous	–
Temperature	Chartevents	36.6	Continuous	[2, 3, 17–19, 22, 23]
Total iron binding capacity	Labevents	320	Continuous	–
Total protein pleural	Labevents	2	Continuous	–
Total bilirubin	Chartevents, labevents	0.7	Continuous	[3, 19, 23]
Total calcium	Labevents	9.6	Continuous	–
Total cholesterol	Labevents	170	continuous	–
Total protein	chartevents, labevents	7	Continuous	–
Triiodothyronine(T3)	Labevents	150	Continuous	–
Uric acid	Chartevents, labevents	4.8	Continuous	–
Vitamin B12	Labevents	230	Continuous	–

The second column is the source table of the variable from the MIMIC-III database; the third column is the value used for missing data imputation in the variable data preprocessing step, which is the normal value selected from the normal range of the corresponding variable; the fourth column describes how the model handles variables



**Fig. 4** Variable correlation heatmap. *CK-MB* creatine kinase MB isoenzyme, *CK* creatinine kinase, *DBP* diastolic blood pressure, *LD* lactate dehydrogenase, *RBC* red blood cell count, *SBP* systolic blood pressure, *TIBC* total iron binding capacity

respectively. Similarly, on the MIMIC test set, the hospital data subset with the same proportion of positive and negative samples as the MIMIC test set, and all the test data, NIMRF-H1 still performs the best, with an AUC of 0.8015 on the three datasets [95% CI 0.7725–0.8304], 0.9115 [95% CI 0.8829–0.9400], 0.8500 [95% CI 0.8303–0.8697].

### Feature importance

For the corresponding memory modules of the four-time window death risk prediction models, the top 25 variables with the highest contribution are shown in Fig. 5a, and there are 21 variables (Table 7) with the highest contribution in no less than 3 models. 25, of which 2 variables ranked in the top 25 in terms of contribution in the 4 models, namely, red blood cell counts and sodium. For the explanation modules of the four models, the top 25 variables with the highest contribution are shown in Fig. 5b. The variables with the top 25 contributions in the four models (Table 8) are the same, and the variables with the top 5 contributions were all the same five vital sign variables: systolic blood pressure, heart rate, diastolic blood pressure, respiratory rate, and body temperature. We have drawn a visual heatmap of the contribution of each variable to the random forest module of the model at each hour (Fig. 6) to show this situation more intuitively. The row represents the data collection time, and the column represents the variable. The brighter the pixel point is, the greater the contribution of the corresponding variable at the corresponding moment. The lit areas in the corresponding heatmaps of the four models are basically the same.

**Table 5** Sensitivity and specificity at different thresholds

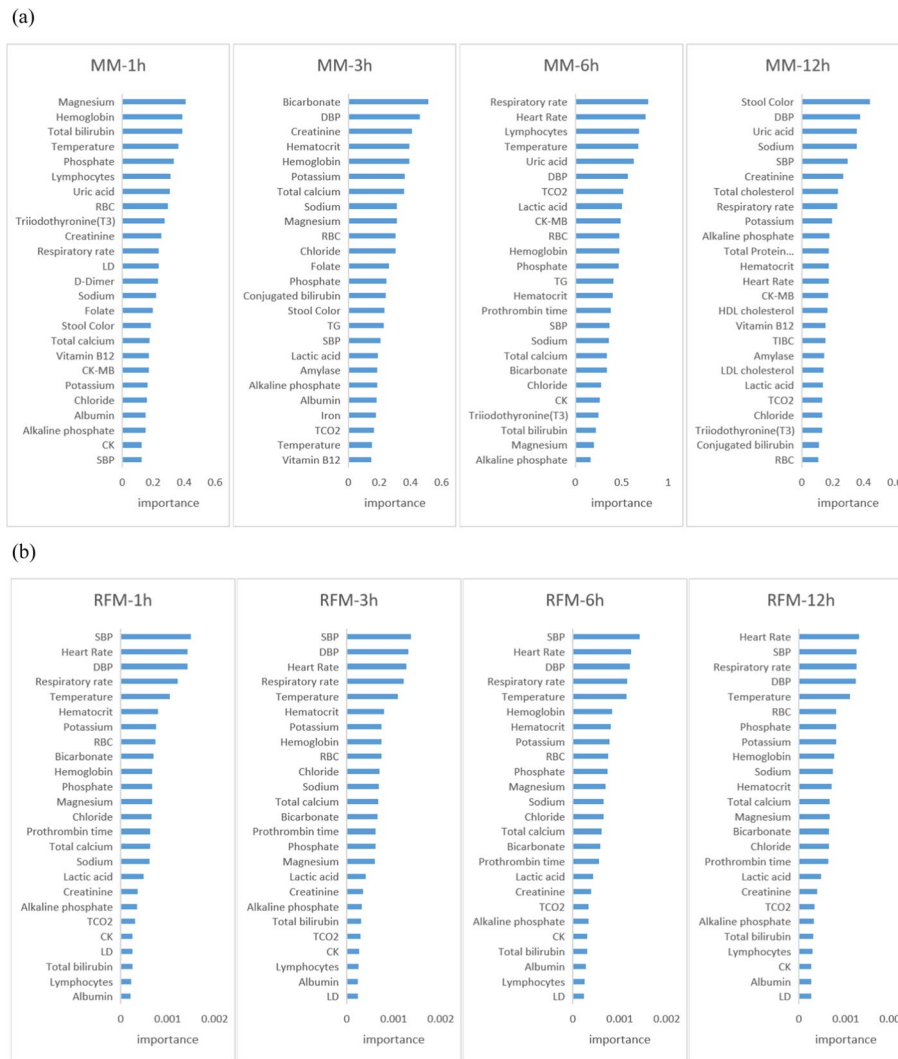
No.	Algorithm	Variables	Model cut-off	Sensitivity (%)	Specificity (%)
1	RF	Model variables	$\geq 0.5692$	51	85
			$\geq 0.5808$	37	90
			$\geq 0.6192$	24	95
			$\geq 0.7009$	9	99
2	LSTM	Model variables	$\geq 0.4298$	66	85
			$\geq 0.4389$	62	90
			$\geq 0.4596$	57	95
			$\geq 0.5622$	41	99
3	Cox	Model variables	$\geq 0.9110$	12	85
			$\geq 0.9162$	7	90
			$\geq 0.9482$	1	95
			$\geq 0.9875$	0	99
4	NIMRF	Model variables	$\geq 0.4179$	89	51
			$\geq 0.4419$	80	74
			$\geq 0.4499$	76	80
			$\geq 0.4588$	72	85
			$\geq 0.4705$	67	90
			$\geq 0.4894$	58	95
5	NIMRF	Reference variables	$\geq 0.5675$	40	99
			$\geq 0.4759$	73	80
			$\geq 0.4885$	69	85
			$\geq 0.5044$	63	90
			$\geq 0.5291$	52	95
6	NIMRF	Model vital sign variables	$\geq 0.5914$	36	99
			$\geq 0.4387$	74	80
			$\geq 0.4461$	70	85
			$\geq 0.4572$	65	90
			$\geq 0.4809$	57	95
			$\geq 0.5694$	39	99

No. 4 is a NIMRF model based on 43 model variables, No. 1, No. 2, and No. 3 are random forest, LSTM, and Cox regression models based on 43 model variables, respectively. No. 5 and No. 6 are NIMRF models based on Reference 18 variables and vital sign variables in our study, respectively

**Table 6** NIMRF model performance (AUC) test

	NIMRF-H1		NIMRF-H3		NIMRF-H6		NIMRF-H12	
	AUC	95% CI	AUC	95% CI	AUC	95% CI	AUC	95% CI
Dataset from MIMIC	0.8015	(0.7725, 0.8304)	0.7100	(0.6777, 0.7423)	0.6761	(0.6425, 0.7097)	0.638	(0.6031, 0.6729)
Dataset from hospital	0.9366	(0.9157, 0.9575)	0.8548	(0.8198, 0.8898)	0.678	(0.6352, 0.7207)	0.6453	(0.6110, 0.6796)
Data subset from hospital	0.9115	(0.8829, 0.9400)	0.7913	(0.7512, 0.8315)	0.6614	(0.6160, 0.7068)	0.6445	(0.5974, 0.6917)
All test data	0.8500	(0.8303, 0.8697)	0.7045	(0.6788, 0.7303)	0.6745	(0.6482, 0.7008)	0.6467	(0.6224, 0.6711)

For the four-time window mortality risk prediction model, the SHAP algorithm can quantify the group factor contribution (global interpretation) and individual factor contribution (local interpretation) of NIMRF. Taking the one-hour mortality risk prediction model as an



**Fig. 5** Ranking of the top 25 variables by contribution

example, the global and local explanations of the model are shown in Figs. 7 and 8. As time series data is used, variables at different times are considered as different features. In our study, there were a total of 4750 related features among the 43 variable indicators. In the feature names in the figure, “value” represents the observed data at a certain time, “mask” represents the observed data at a certain time, and the subscript number represents the time. Figure 7a shows that for the top 50 important features in predicting the risk of death within one hour, there are a total of 18 vital sign related features, involving all vital sign variables included in our study, namely systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate, and body temperature. The blue color in Fig. 7b indicates a small observed value of the characteristic factor, while the red color indicates a large observed value of the characteristic factor. The horizontal axis represents the SHAP value. Generally, the higher the SHAP value, the higher the risk of death. Figure 7b shows that: (1) for a certain moment, the SBP for multiple hours before that moment is negatively correlated with the risk of death within one hour after that moment, listed in descending order of feature importance as the previous 7th, 2nd, 25th, and



**Table 7** List of top 25 variables in NIMRF memory module contribution

Variable	Rank of the importance			
	NIMRF-H1 (MM)	NIMRF-H3 (MM)	NIMRF-H6 (MM)	NIMRF-H12 (MM)
Magnesium	1	9	24	28
Hemoglobin	2	5	11	36
Temperature	4	24	4	30
Phosphate	5	13	12	43
Uric acid	7	43	5	3
Red blood cell count	8	10	10	25
Triiodothyronine (T3)	9	35	22	23
Creatinine	10	3	42	6
Respiratory rate	11	34	1	8
Sodium	14	8	17	4
Stool color	16	15	34	1
Total calcium	17	7	18	31
Vitamin B12	18	25	43	16
Creatine kinase MB isoenzyme	19	36	9	14
Potassium	20	6	29	9
Systolic blood pressure	25	17	16	5
Diastolic blood pressure	33	2	6	2
Lactic acid	40	18	8	20
Hematocrit	42	4	14	12

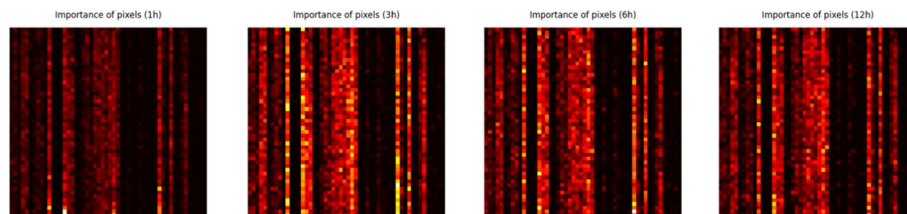
The contribution is ranked in the top 25 in the memory module of no less than 3 models

24th hours (systolic blood pressure value 42, systolic blood pressure value 47, systolic blood pressure value 24, systolic blood pressure value 25); (2) In addition, the diastolic blood pressure (diastolic blood pressure value 47) in the second hour prior to a certain moment, heart rate (heart rate value 49, heart rate value 48) in the first hour, respiratory rate (respiratory rate value 49) in the first hour, and body temperature (temperature value 13) in the previous 36 h are negatively correlated with the risk of death within the first hour after that moment.

Figure 8 shows the individual factor contribution values corresponding to randomly selected samples of two different patients, with red indicating unfavorable factors and blue indicating favorable factors. Figure 8a shows that the systolic blood pressure and diastolic blood pressure of the patient are unfavorable factors at multiple times, and their blood pressure has initially increased in the past 48 h, and then continued to decrease within 9 h before death (systolic blood pressure value 12, 42, 46, 47 were 0.466, 0.545, 0.466, 0.388, respectively, and diastolic blood pressure value 6, 40, 46, 47 were 0.413, 0.5, 0.423, 0.365, respectively). The sum of the contribution values of all factors plus the baseline value of 0.519 is  $0.692 > 0.5$ , indicating that the patient died within one hour after that time, which is consistent with the fact. Figure 8b shows that the important unfavorable factors of the patient are laboratory variables at certain times [Bicarbonate, Prothrombin, Chlorine, Creatinine Kinase (CK), Lactic acid, Red blood cell count], and the important favorable factors include multiple vital sign variables (systolic blood pressure, diastolic blood pressure, heart rate). The total contribution value of all factors plus the baseline value of 0.519 is  $0.385 < 0.5$ , so the patient will not die within one hour after that moment, which is consistent with the fact.

**Table 8** List of the top 25 variables in the NIMRF random forest module contribution

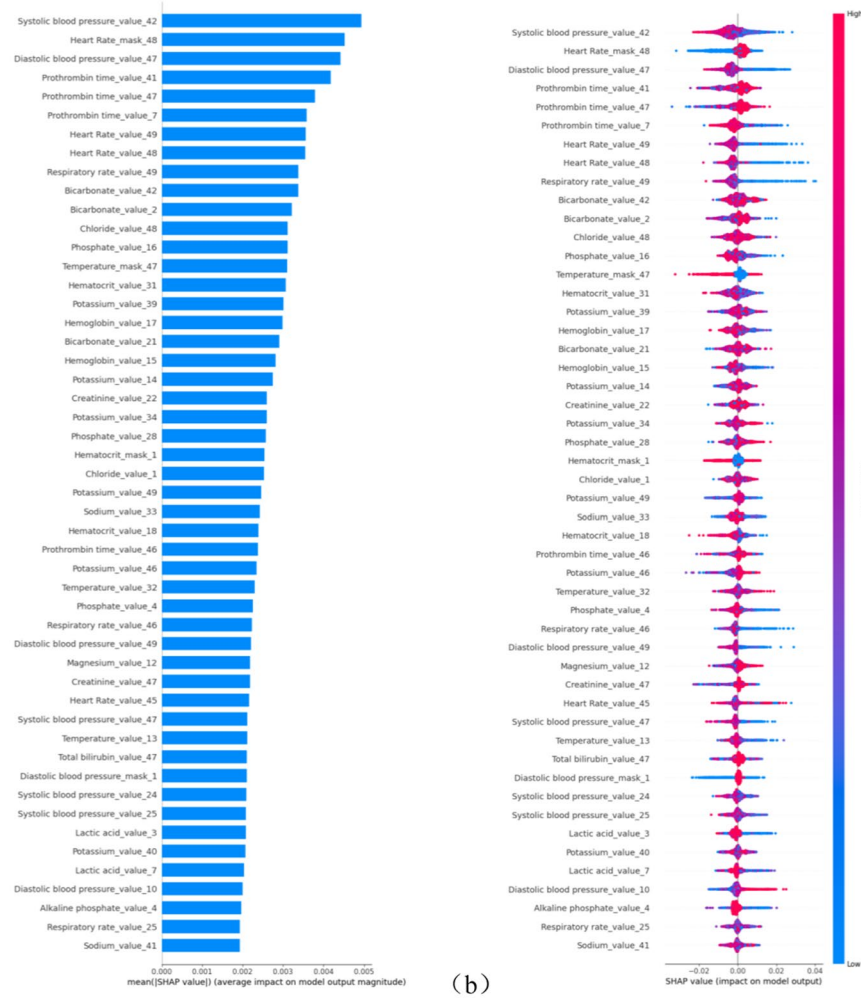
Variable	Rank of the importance			
	NIMRF-H1 (MM)	NIMRF-H3 (MM)	NIMRF-H6 (MM)	NIMRF-H12 (MM)
Systolic blood pressure	1	1	1	2
Heart rate	2	3	2	1
Diastolic blood pressure	3	2	3	4
Respiratory rate	4	4	4	3
Temperature	5	5	5	5
Hematocrit	6	6	7	11
Potassium	7	7	8	8
Red blood cell count	8	9	9	6
Bicarbonate	9	13	15	14
Hemoglobin	10	8	6	9
Phosphate	11	15	10	7
Magnesium	12	16	11	13
Chloride	13	10	13	15
Prothrombin time	14	14	16	16
Total calcium	15	12	14	12
Sodium	16	11	12	10
Lactic acid	17	17	17	17
Creatinine	18	18	18	18
Alkaline phosphate	19	19	20	20
TCO <sub>2</sub>	20	21	19	19
Creatinine kinase (CK)	21	22	21	23
Lactate dehydrogenase	22	25	25	25
Total bilirubin	23	20	22	21
Lymphocytes	24	23	24	22
Albumin	25	24	23	24



**Fig. 6** Heatmap of the variable contribution of the NIMRF random forest module. The row represents the time of the sample data, and the column represents the variable; the brighter the pixel, the greater the contribution of the corresponding variable at the corresponding moment. The lighted areas in the corresponding heatmaps of the four models are basically the same, indicating that the random forest module for the four models is the variables with high contribution are the same

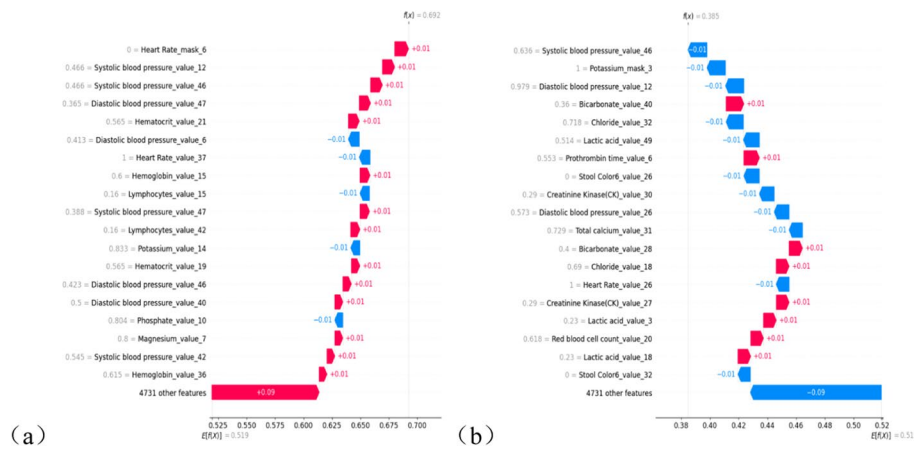
## Discussion

In this study, we developed the NIMRF model to predict the mortality risk of ICU patients in real-time based on multidimensional data on vital sign variables and laboratory variables in the MIMIC-III database, explained the model, and found that vital signs are important for patient prognosis. Predictive value, in addition to external validation of the model using data from the Chinese PLA General Hospital.



**Fig. 7** Global interpretation of the one hour mortality risk prediction model. The vertical axis represents the characteristic factor, where "value" represents the observed data at a certain moment, "mask" represents the observed data at a certain moment, and the subscript number represents the moment. The horizontal axis of Figure (a) represents the result of taking the absolute value of SHAP in Figure (b) first and then taking the average value. The horizontal axis in Figure (b) represents the SHAP value, where each point in Figure (b) represents a sample, with blue indicating small feature values and red indicating large feature values

Most of the prediction models for the death of ICU patients are based on data from the initial period of admission to predict 28-day mortality or in-hospital mortality [17, 26]. This predictive model is difficult to provide timely information and has limited impact on daily evaluation and treatment. In this study, four models were developed based on the developed NIMRF network to predict the mortality risk of patients in four nonoverlapping time windows of 0–1 h, 1–3 h, 3–6 h, and 6–12 h after any time. In this way, real-time dynamic monitoring of the condition of ICU patients can be performed to provide more information. Based on the hospital dataset, the AUCs of the four NIMRF models for predicting the risk of death in patients in the aforementioned four-time windows were 0.9366 [95% CI 0.9157–0.9575], 0.8548 [95% CI 0.8198–0.8898], 0.6780 [95% CI 0.6352–0.7207], 0.6453 [95% CI 0.6110–0.6796].



**Fig. 8** Partial explanation of the one hour mortality risk prediction model. The horizontal axis represents the SHAP value, and the vertical axis represents the feature. Red indicates that features have a positive impact on prediction (arrow to the right), while blue indicates that features have a negative impact on prediction (arrow to the left). The  $E[f(x)]$  below is the baseline value of SHAP, and the  $f(x)$  above is the total SHAP value of the sample

Overall, we can now predict with high accuracy the risk of death across time windows after any given moment in time.

Considering the high complexity of the task of predicting the prognosis of ICU patients, we based our current study on four test datasets (MIMIC test set, hospital dataset, hospital data subset, and the full composition of the MIMIC test set and hospital dataset) to evaluate three different prediction methods: NIMRF, LSTM and random forest. The result is that for the prediction of death risk in different time windows of different test sets, the NIMRF model predicts the highest AUC value, and the sensitivity of NIMRF is higher than the other two under the same specificity, which indicates that the NIMRF model currently developed by us has the best prediction performance and model availability. In addition, although the MIMIC test set and the hospital data subset have similar population characteristics, the test performance on the hospital data subset is higher than that on the MIMIC test set, which may be related to the more realistic and reliable death time of the hospital data.

The ranking of the contribution of the feature variables indicates the importance of each variable to the prediction. In this study, for the random forest modules of the four models, the top 5 contributing variables are all vital sign variables (systolic blood pressure, heart rate, diastolic blood pressure, respiratory rate and body temperature), which can be understood that the vital signs are more essential for model prediction. In actual clinical practice, vital signs are important signs that mark the existence and quality of life activities and are one of the important items for evaluating the body. Monitoring the vital signs of patients in the intensive care unit (ICU) is absolutely necessary to help evaluate overall health [27]. There are currently some big data studies based on vital signs. Daniel et al. [27] used machine learning to predict heart rate, blood oxygen level ( $SpO_2$ ), mean arterial pressure (MAP), respiratory rate (RR) and systolic blood pressure (SBP) in the next hour. Mohamadlou et al. [28] used a gradient boosting algorithm to predict severe AKI characterized by vital signs and Scr, this

algorithm achieved AUROC 0.872 during the onset of the disease. For the prediction of 12 h, 24 h, 48 h, and 72 h before onset, the algorithm achieved AUROC values of 0.800, 0.795, 0.761, and 0.728, respectively. Sivasubramaniam et al. [29] validated the feasibility of sepsis classification based on vital sign data. Heart rate, respiratory rate and blood pressure can be used as important classification value, and fluid therapy for blood pressure can also significantly change the prognosis of patients. This study also shows that vital signs are an important factor affecting the prognosis of ICU patients, and clinicians need to pay more attention to the management and regulation of vital signs with higher contribution.

Our research is based on the MIMIC-III dataset to develop a model, including multi-dimensional features such as vital sign variables and laboratory variables, and externally validated findings related to hospital data, with the aim of predicting the risk of short-term mortality in real time and assisting clinicians in making timely decisions about treatment and intervention. There are some limitations in our study. First, we choose the observation data window length of 48 h, and the time windows are 0–1 h, 1–3 h, 3–6 h, and 6–12 h. This setting can be used as an example to demonstrate the feasibility of real-time dynamic prediction of patient mortality risk in ICU based on machine learning. In practical applications, the corresponding observation data window duration and prediction time window can be set according to clinical needs. Second, the current model has the best performance in predicting the risk of death within 1 h, and clinical decisions can be made with greater reference to predictive information from this time window; with the extension of time, the treatment measures and treatment background will change. Data heterogeneity causes the performance of the prediction model to decrease with the delay of the prediction window (1–3 h, 3–6 h, 6–12 h), but the impact can be minimized by expanding the training data set, increasing the diversity of the training samples, and attempting to model clinical intervention variables such as treatment. Third, from the explanatory results of the model in our study, vital sign variables are very important for predicting the mortality risk of ICU patients, and in practical applications, vital sign variables are characterized by a large amount of data, high reliability, and ease of processing and analysis. In the study of ICU patients, an attempt can be made to build a mortality risk prediction model for ICU patients by filtering data only for vital sign variables. Finally, Goh et al. show that unstructured data can improve the performance of the model [30], and we will consider adding some unstructured data features to the feature set.

## Conclusion

In this study, based on multidimensional variables such as vital sign variables and laboratory variables, we used machine learning to establish a dynamic prediction model to predict the mortality risk of ICU patients in four nonoverlapping time windows with high accuracy. The interpretation of the model proposes factors that affect the prediction. These predictions can assist clinicians in choosing more timely and reasonable treatments and interventions. Based on ICU multisource heterogeneous data, this study established a real-time dynamic mortality risk model using machine learning algorithms. The NIMRF method may provide more stable and reliable results, especially at the first hour of judgment. Data on vital signs showed encouraging predictive effects and

became an important indicator of prognosis. That is, with the help of big data modeling and methodology, these easily accessible vital sign parameter data have gained a new lease of life, showing the indicator advantages of continuous and dynamic analysis. This can provide a prognostic method that can be easily generalized and applied to assist clinicians in selecting more timely and rational treatments and interventions.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40537-024-00896-8>.

**Additional file 1.** Proportional risk hypothesis test.

## Acknowledgements

The authors thank the president of Shandong Future Network Research Institute and Jiangsu Future Network Group Co for their continuous encouragement and support.

## Author contributions

All mentioned authors contribute to the elaboration of the paper. PP and YW wrote the main manuscript text; YT and HC prepared figures and tables. CL and QY conducted statistical analysis. FX, YL, LX, YL were responsible for revising and reviewing articles. All authors reviewed the manuscript. All authors read and approved the final manuscript.

## Funding

This research was supported by Logistics research independent innovation project; the Chinese PLA General Hospital Youth Independent Innovation Research Project (No. 22QNFC146); the key project of the Eighth Medical Center of Chinese PLA General Hospital (No. 2021ZD001).

## Availability of data and materials

All data used is publicly available.

## Declarations

### Ethics approval and consent to participate

Yes, consent is granted.

### Consent for publication

Yes, consent is granted for publication.

### Competing interests

The authors declare that they have no competing interests.

Received: 1 June 2023 Accepted: 19 February 2024

Published online: 18 April 2024

## References

1. Apgar V. A proposal for a new method of evaluation of the newborn. *Curr Res Anesth Analg*. 1952;32:260–7.
2. Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE-acute physiology and chronic health evaluation: a physiologically classification system. *Crit Care Med*. 1981;9:591.
3. Le Gall J-R, et al. A simplified acute physiology score for ICU patients. *Crit Care Med*. 1984;12:975–7.
4. Apolone G, Bertolini G, et al. The performance of SAPS II in a cohort of patients admitted to 99 Italian ICUs: results from GIVITI. *Intensive Care Med*. 1996;22(12):1368–78.
5. Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicenter, Portuguese study. *Intensive Care Med*. 1997;23(2):177–86.
6. Rui M, Miranda DR, Fidler V, et al. Evaluation of two outcome prediction models on an independent database. *Crit Care Med*. 1998;26(1):50–61.
7. Rowan KM, Kerr JH, Major E, et al. Intensive Care Society's APACHE II study in Britain and Ireland—II: outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. *BMJ Clin Res*. 1993;307(6910):977–81.
8. Bastos PG, Sun X, Wagner DP, et al. Application of the APACHE III prognostic system in Brazilian intensive care units: a prospective multicenter study. *Intensive Care Med*. 1996;22:564–70.
9. Zimmerman JE, Wagner DP, Draper EA, et al. Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. *Crit Care Med*. 1998;26(8):1317–26.
10. Popovich MJ. If most intensive care units are graduating with honors, is it genuine quality or grade inflation? *Crit Care Med*. 2002;30(9):2145–6.
11. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med*. 2006;34(5):1297–310.

12. Le Gall JR, Neumann A, Hemery F, Bleriot JP, Fulgencio JP, Garrigues B, Gouzes C, Lepage E, Moine P, Villers D. Mortality prediction using SAPS II: an update for French intensive care units. *Crit Care*. 2005;9(6):R645–52.
13. Ray S. A quick review of machine learning algorithms. In: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). 2019.
14. Awad A, et al. Predicting hospital mortality for intensive care unit patients: time-series analysis. *Health Inform J*. 2019;26(2):1043–59.
15. Gong JJ, Naumann T, Szolovits P, Guttag JV. Predicting clinical outcomes across changing electronic health record systems. In: Proceedings of KDD'17, Halifax, NS, Canada. 2017. p. 9.
16. Rajkomar A, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med*. 2018;1:18.
17. Ghassemi M, Naumann T, Doshivelez F, et al. Unfolding physiological state: mortality modelling in intensive care units. In: ACM SIGKDD international conference on knowledge discovery & data mining. KDD; 2014.
18. Zhang G, Xu JM, Yu M, et al. A machine learning approach for mortality prediction only using non-invasive parameters. *Med Biol Eng Comput*. 2020;50:2195–238.
19. Kim SY, Kim S, Cho J, Kim YS, Sol IS, Sung Y, et al. A deep learning model for real-time mortality prediction in critically ill children. *Crit Care*. 2019;23(1):1–10.
20. Celi LA, Hinske LC, Alterovitz G, et al. An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study. *Crit Care*. 2008;12(6):1–17.
21. Hu J, Kang XH, Xu FF, Huang KZ, Du B, Weng L. Dynamic prediction of life-threatening events for patients in intensive care unit. *BMC Med Inform Decis Mak*. 2022;22(1):276.
22. Allen D. Automatic one-hot re-encoding for FPLs. In: Selected papers from the second international workshop on field-programmable logic and applications, FieldProgrammable gate arrays: architectures and tools for rapid prototyping. London: Springer-Verlag. 1993. p. 71–7.
23. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–80.
24. Breiman L. Random forest. *Mach Learn*. 2001;45:5–32.
25. Cox DR. Regression models and life tables. *J R Stat Soc Ser B*. 1972;34(2):187–220.
26. Citi L, Barbieri R. PhysioNet 2012 challenge: predicting mortality of ICU patients using a cascaded SVM-GLM paradigm. In: Computing in cardiology; IEEE. 2013.
27. Chang D, Chang D, Pourhomayoun M. Risk prediction of critical vital signs for ICU patients using recurrent neural network. In: 2019 international conference on computational science and computational intelligence (CSCI 2019), 5–7 Dec. 2019, Las Vegas, NV, USA. 2019. p. 1003–6.
28. Mohamadlou H, Lynn-Palevsky A, Barton C, et al. Prediction of acute kidney injury with a machine learning algorithm using electronic health record data. *Can J Kidney Health Dis*. 2018;5:1–9.
29. Bhavani SV, Semler M, Qian ET, Verhoef PA, Robichaux C, Churpek MM, Coopersmith CM. Development and validation of novel sepsis subphenotypes using trajectories of vital signs. *Intensive Care Med*. 2022;48(11):1582–92.
30. Goh KH, Wang L, Yeow A, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun*. 2021;12(1):1–10.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.