

RESEARCH

Open Access



Dynamic prediction of life-threatening events for patients in intensive care unit

Jiang Hu^{1,2†}, Xiao-hui Kang^{1†}, Fang-fang Xu², Ke-zhi Huang², Bin Du¹ and Li Weng^{1*}

Abstract

Background: Early prediction of patients' deterioration is helpful in early intervention for patients at greater risk of deterioration in Intensive Care Unit (ICU). This study aims to apply machine learning approaches to heterogeneous clinical data for predicting life-threatening events of patients in ICU.

Methods: We collected clinical data from a total of 3151 patients admitted to the Medical Intensive Care Unit of Peking Union Medical College Hospital in China from January 1st, 2014, to October 1st, 2019. After excluding the patients who were under 18 years old or stayed less than 24 h at the ICU, a total of 2170 patients were enrolled in this study. Multiple machine learning approaches were utilized to predict life-threatening events (i.e., death) in seven 24-h windows (day 1 to day 7) and their performance was compared.

Results: Light Gradient Boosting Machine showed the best performance. We found that life-threatening events during the short-term windows can be better predicted than those in the medium-term windows. For example, death in 24 h can be predicted with an Area Under Curve of 0.905. Features like infusion pump related fluid input were highly related to life-threatening events. Furthermore, the prediction power of static features such as age and cardio-pulmonary function increased with the extended prediction window.

Conclusion: This study demonstrates that the integration of machine learning approaches and large-scale high-quality clinical data in ICU could accurately predict life-threatening events for ICU patients for early intervention.

Keywords: Early prediction, Deterioration, Life-threatening events, Mortality, Machine learning

Background

Critically ill patients are admitted to Intensive Care Unit (ICU) for constant care and continuous organ support [1]. Frequent lab tests and medications are provided until patients are discharged. Patients who are admitted to ICU are very sick and therefore they tend to have a higher mortality rate than the average patients on the wards. With the development of artificial intelligence, many

researchers have applied machine learning methods to study severe illnesses, such as sepsis [2–7], Acute kidney injury (AKI) [8–11], hypotension [12–14], etc. Early prediction of patients at high risk of deterioration within a short period of time would trigger immediate attention and intervention and thus would reduce the mortality rate at ICU, especially in low- and middle-income countries with limited health care resources [15–17].

Machine learning (ML) is a subfield of artificial intelligence that enables a computer to learn from data to make predictions, and the learning performance improves with more training data [18]. A variety of ML methods such as clustering analysis and pattern classification have been used in medical science. Clustering analysis is used to discover cluster structures within data in an unsupervised manner while pattern classification is based

[†]Jiang Hu and Xiao-hui Kang Contributed equally to the work

*Correspondence: wengli@gmail.com

¹ Medical Intensive Care Unit, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, 1 Shuai Fu Yuan, Beijing 100730, China

Full list of author information is available at the end of the article



on labeled data (i.e., supervised learning). ML techniques can more effectively reduce errors and expenses compared with manual operation [19] and have better performance in prediction than traditional scoring methods, such as APACHE, SAPS, and SOFA [17]. Several studies were carried out to predict patient outcomes in ICU. For example, Gong et al. applied a logistic regression approach to predict patient outcomes based on the features extracted from the free-text notes in the Electronic Health Records (EHR) [20]; Johnson et al. predicted patient outcomes using a Bayesian Ensemble method [21]. However, these studies only included data within 72 h of ICU admission while excluding critically ill patients who stayed at ICU for over 3 days. Moreover, most machine learning models for ICU patients utilized a small number of features to predict the risk of mortality [22]. Given the high heterogeneity of patients at ICU, a limited number of features may not be able to capture some important signs of deterioration, leading to sub-optimal performance. To better understand a prediction model and feature contributions, interpretable models were often utilized [23, 24]. For example, MAP, bicarbonate, and creatinine were found to be the top three important features contributing to predict acute gastrointestinal bleed [24].

So far, most of the published machine learning studies of ICU patients at ICU were based on an open database, Medical Information Mart for Intensive Care (MIMIC), which was established by MIT Institute of Computational Physiology, Beth Israel Deaconess Medical Center (BIDMC) and Philips Medical [25]. Few studies were carried out based on ICU data from China.

This study aims to develop a dynamic model for predicting life-threatening events, based on the data of the patients admitted to the Medical Intensive Care Unit of Peking Union Medical College Hospital (PUMCH-MICU).

Methods

Study population and data set

This study was based on the database of the Medical ICU in PUMCH. The dataset includes 3151 patients admitted to the PUMCH-MICU from January 1st, 2014, to October 1st, 2019. The patients who were less than 18 years old or stayed at ICU for less than 24 h were excluded. Five-fold cross-validation was employed to select samples for training and testing, i.e., the ratio of the sample size in a training set to a testing set is 4. The target events of our prediction included not only the outcome of mortality, as most similar studies did, but also predicted life-threatening events which happened during hospitalization. All cardiac arrest-related cardiopulmonary resuscitation was considered life-threatening event [26]. Therefore, in

this study we defined life-threatening event as any signs of cardiac arrest, i.e., the operations of electric defibrillation or chest compression. Because palliative care was not an indication for ICU admission, all decedents received either electric defibrillation or chest compression. We used life-threatening event instead of cardiac arrest because chest compression usually started when the heart rate was declining. The data of vital signs and mechanical ventilation were automatically transferred to the database at intervals of 15 s and 1 min, respectively. The value of laboratory findings including blood gases were automatically collected. The inputs and outputs were manually recorded by registered nurses every hour. In this study, we collected data every hour on the hour and instant data at the life-threatening moments, which were manually confirmed by registered nurses. In addition, a statistic analysis was performed to detect the outliers. If a value was beyond the four standard deviations of the mean, the data would not be considered. In this study, the mean and standard deviation of the percentage of the feature value outliers in the whole cohort were 0.43% and 0.0047, respectively, suggesting high quality of the collected data.

Feature selection

Figure 1 illustrates our study design. A 24-h observational data window for each patient was used and the outcomes in the subsequent 7 days were determined. About 100 features (variables) were selected based on feature importance score computed by an embedded method Random Forest, and the features with importance score equal to zero were removed. The selected features were categorized into two groups, i.e., static features and dynamic features. Table 1 shows demographic features, vitals, lab testing results, hemodynamics, blood gases, inputs, outputs, and ventilator measurements. During each data window, vital, input and output features from each patient were collected hourly, and the tests for hemodynamic, lab, blood gas were usually measured once every several hours. Measurements of ventilators were only collected when available. For dynamic features, different methods were implemented to extract features. For hemodynamic, blood gas and vital features, and derivative features such as mean, max and min were computed. For each feature in the input and output categories, the sum of its values over 24 h was calculated.

Imputation of missing values

In this study, different imputation approaches were applied for different types of variables. The mean imputation was used for demographic variables, and a sample and hold method was used for imputing variables of vital, blood gas, hemodynamic and lab.

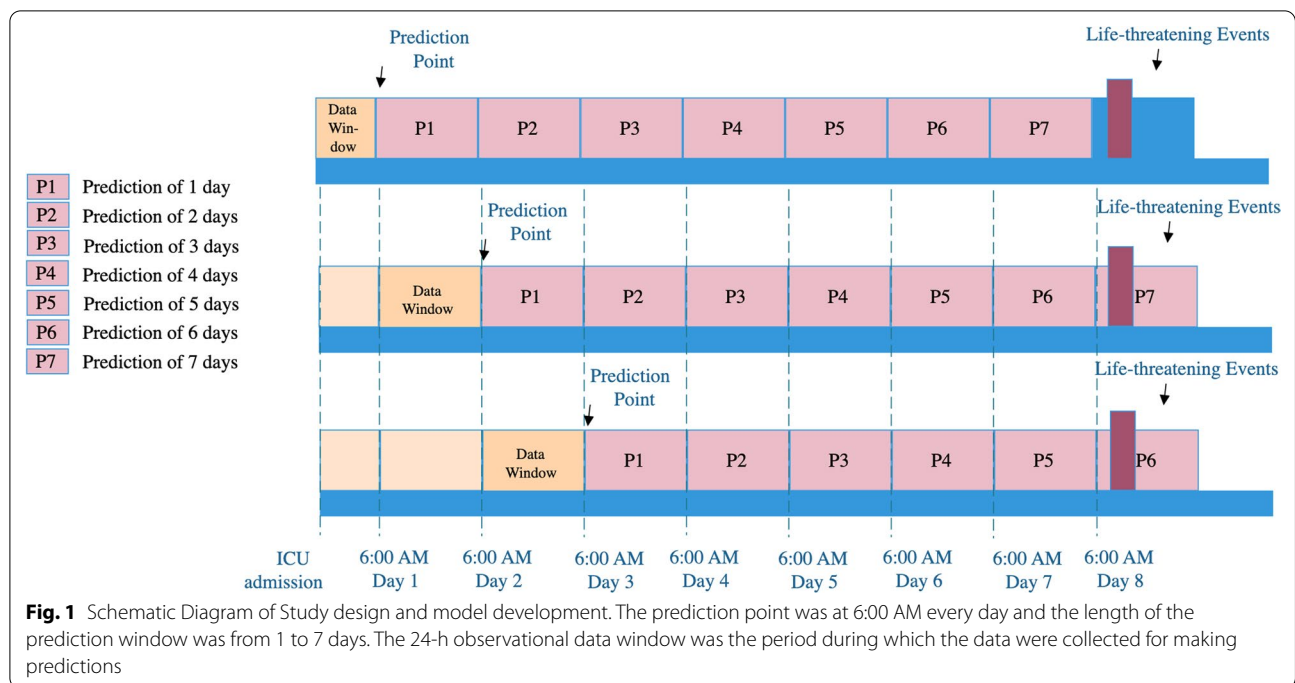


Table 1 Static and dynamic features used in this study. These features are classified into eight categories

Category	Features (Variables)	Derivative Features	Static/dynamic
Demographic	Gender, cardiorespiratory function, CPR, Admission type, ARF, Shock, Coronary Artery Disease, Diabetes, Hypertension, COPD, NYHA class IV, Solid tumors(metastases), Lymphoma, Leukemia, ILD, dialysis, cirrhosis, hepatic failure, myeloma, Rheumatic immune disease, glucocorticoid, immunologic deficiency		Static
Vital	HR, SPO ₂ , ADBP, ASBP, AMBP, Pulse, Resp, Temper, NDBP, NSBP, Glasgow coma scale (GCS), GCS Eye opening, GCS Verbal Response, GCS Motor Response, RASS, PAPm, CVP	Mean, Max, Min	Dynamic
Hemodynamic	SVRI, CI, BSA, SVR(Picco), SI(Picco), LVSW	Mean, max, Min	Dynamic
Blood Gas	pH, Ca, Cl, K, BE, Na, Lac, HCO ₃ , tHb, AnionGap, Hct, sO ₂ , pCO ₂ , O ₂ Hb, COHb, HHb, MetHb, Glu, pO ₂ , PaO ₂ -FIO ₂	Mean, Max, Min	Dynamic
Lab	HGB, WBC, RBC, MCV, HCT, PLT, EOS_PERCENT, LY_PERCENT, NEUT_PERCENT, MONO_ABS, Creatinine, MONO_PERCENT, RDW_S, PCT, Ca, TCO ₂ , Urea, PDW, Glu, TBil, ALT, DBil, INR, PT	Mean, Max, Min	Dynamic
Ventilator	FiO ₂ _Set, VT _e , MV _e , RR, P _{peak} , PS, RR_Set	Mean	Dynamic
Input	Injection, Nasal Feeding, Intravenous Intake, Infusion pump related fluid input, Norepinephrine, Midazolam, Furosemide, Epinephrine	Sum	Dynamic
Output	Urine Volume, Stool count	Sum	Dynamic

CPR cardiac pulmonary resuscitation, *ARF* acute renal failure, *COPD* chronic obstructive pulmonary disease, *ILD* interstitial lung disease, *HR* heart rate, *SPO₂* pulse oximeter oxygen saturation, *ADBP* arterial diastolic blood pressure, *ASBP* arterial systolic blood pressure, *AMBP* arterial mean blood pressure, *NDBP* noninvasive diastolic blood pressure, *NSBP* noninvasive systolic blood pressure, *NMBP* noninvasive mean blood pressure, *GCS* Glasgow Coma Scale, *RASS* Richmond Agitation–Sedation Scale, *PAP* pulmonary artery pressure, *CVP* central venous pressure, *SVRI* systemic vascular resistance index, *CI* cardiac index, *PICCO* pulse indicator continuous cardiac output, *P_{SV}* PICCO_stroke volume, *BSA* body surface area, *P_{SVR}* PICCO systemic vascular resistance, *P_{SI}* PICCO stroke index, *LVSW* left ventricular stroke work, *PCO₂* pressure of carbon dioxide, *Ca* calcium, *Cl* chloride, *K* potassium, *BE* base excess, *Na* sodium, *Lac* lactate, *tHb* total hemoglobin, *Hct* hematocrit, *sO₂* Oxygen saturation, *pCO₂* = pressure of carbon dioxide, *O₂Hb* oxyhemoglobin, *COHb* carboxhemoglobin, *HHb* deoxyhemoglobin, *MetHb* methemoglobin, *Glu* glucose, *pO₂* pressure of oxygen, *PaO₂* arterial oxygen pressure, *FIO₂* fractional inspired oxygen, *HGB* hemoglobin, *WBC* white blood cell, *MCV* mean corpuscular volume, *HCT* hematocrit, *PLT* platelet, *EOS_PERCENT* percentage of eosinophil, *LY_PERCENT* percentage of lymphocyte, *NEUT_PERCENT* percentage of neutrophil, *MONO_ABS* monocytes absolute, *RDW* red cell distribution width, *PCT* procalcitonin, *TCO₂* total carbon dioxide, *PDW* platelet distribution width, *TBil* total bilirubin, *ALT* alanine aminotransferase, *DBil* bilirubin, *INR* international normalized ratio, *PT* prothrombin time, *VT_e* tidal volume expiration, *MV_e* minute volume expiration, *RR* respiratory rate, *P_{peak}* peak pressure, *PS* pressure support, *RR_{set}* the set of respiratory_rate

Model development

The model construction was shown in Fig. 1. At each prediction time point at 6:00AM of each day, seven

predictors were developed to predict the life-threatening events. Four machine learning methods including Logistic Regression (LR), Support Vector Machine

(SVM), Random Forest (RF), and light Gradient Boosting Machine (lightGBM) were employed to perform the prediction. LR is a generalized linear regression model and has a good fit for linear relationship between predictor and outcome variables [18]. SVM is a well-known machine learning algorithm for linear or nonlinear classification by choosing different kernel functions [18]. RF is an ensemble algorithm that consists of many weak learners and outputs the final decision by majority voting strategy or mean strategy [27]. lightGBM is a novel machine-learning algorithm with excellent performance. As a combination of tree-based learning algorithms and a gradient boosting framework, lightGBM is capable of handling large-scale data [28–30] and achieves good performance with small data as well [31]. The boosting algorithm is based on a set of weak learners. The residual loss function is minimized through the sequential trees. In addition, at the training stage a repeated k-fold cross-validation was used. At the testing stage, the combined estimator predicted the probability of life-threatening events by a majority voting strategy.

Statistical analysis

The characteristics of static features in the training and testing sets are shown in Table 2. The continuous variables are presented as mean \pm SD. The performance of the ensemble method was estimated by the metrics of Area Under the Receiver Operating Curve (AUROC/AUC). The 95% confidence intervals (CI) of AUC were calculated by bootstrapping. In addition, sensitivity, specificity, and positive predictive value (PPV) were calculated under the proper cutoff point determined by Youden's index [32]:

$$YI = \text{sensitivity} + \text{specificity} - 1$$

A cutoff point is determined when YI reaches the maximum.

Results

A total of 2170 patients admitted at PUMCH-MICU were included in the final analysis, as shown in Fig. 2. Most of the patients were admitted from the Department of Emergency Internal Medicine and 333 (15%) of the patients had life-threatening events during the study period.

Characteristics of the static variables in the PUMCH-MICU dataset

Table 2 shows the characteristics of the static variables. The mean ages in the training and testing sets were 56(\pm 19) and 57(\pm 18), respectively.

Determination of repeated k-fold cross-validation

As shown in Fig. 1, seven predictors were constructed to predict outcomes at different lengths of periods (i.e., day 1 to day 7) and were labeled as GBM-D1, GBM2, ..., and GBM-D7. To find the best repeated times for k-fold cross-validation, we used GBM-D1 to test the performance with a range between 1 and 15 for repetitions. The results indicated that 4 repetitions for fivefold cross-validation achieved the best performance (Fig. 3).

The performance comparison of four methods

In this study, four different prediction methods were implemented, and their prediction performance was compared, as shown in Fig. 4. Both LR and SVM have an AUC of about 84% while RF and GBM had better performance (AUCs > 90%). Among the four methods, GBM had the best performance with an AUC of 90.5% on day 1. Therefore, we focused more on GBM in the subsequent analyses.

Model development of life-threatening prediction

Finally, the performance of the seven predictors (GBM-D1, ..., GBM-D7) on the testing data is shown in Table 3. As expected, GBM-D1 has the best performance with an AUC of 0.905 [95% CI 0.863, 0.930]). The AUC of GBM-D2 drops to 0.887 (95% CI 0.877, 0.922). The overall performance decreases gradually from 90.5 to 87.2% with the days increasing from 1 to 7. While GBM-D7 has the lowest AUC of 0.872 (95% CI 0.853, 0.887), its PPV of 0.231 is the highest (95% CI 0.195, 0.263).

Feature importance

As the lightGBM algorithm is based on tree-based weaker learners, it offers interpretable models. The top 25 most important features were shown in Fig. 5. Lac, Infusion pump related fluid input and CVP were the most important features in all the models. The ranking of norepinephrine decreases when the prediction window extends from day-1 to day-7. We also observed that the rankings of static features such as age, cardio-pulmonary function, and lymphoma increased as the prediction window extended.

Sensitivity analysis

Considering that atrial fibrillation is not always life-threatening, defining electric defibrillation as a life-threatening event might not be accurate. As such, we performed a sensitivity analysis by excluding all the cases with electric defibrillation in the classification system. After excluding 29 cases of electric defibrillation, we found the main findings of our study remained

Table 2 Characteristics of the static features in the dataset

	Training		Testing	
	Patients without life-threatening events	Patients with life-threatening events	Patients without life-threatening events	Patients with life-threatening events
	n = 1496	n = 274	n = 341	n = 59
Age (year, mean \pm SD)	55 \pm 19	62 \pm 17	57 \pm 18	59 \pm 19
APACHE II score (mean \pm SD)	19 \pm 7	25 \pm 8	18 \pm 6	23 \pm 6
<i>Gender</i>				
Female	702 (47)	117 (43)	140 (41)	21 (36)
Male	794 (53)	157 (57)	201 (59)	38 (64)
CPR	36(4)	15(9)	10(3)	4(7)
<i>Admission Type</i>				
Emergency non-operation	677 (81)	127 (93)	226 (78)	46 (92)
Emergency Operation	91 (11)	5 (4)	41 (14)	4 (8)
Elective non-operation	40 (5)	3 (2)	9 (3)	0 (0)
Elective Operation	23 (3)	1 (1)	15 (5)	0 (0)
Elective transferred	120	31	38	9
<i>Complications</i>				
ARF	192 (13)	70 (26)	37 (11)	10 (17)
Shock	503 (37)	169 (64)	160 (48)	42 (72)
<i>Comorbidities</i>				
Coronary Artery Disease	198 (13)	54 (20)	46 (13)	14 (24)
Diabetes	224 (15)	59 (22)	63 (18)	9 (15)
Hypertension	438 (29)	98 (36)	128 (38)	17 (29)
COPD	66 (4)	15 (5)	6 (2)	1 (2)
NYHA class IV	61 (4)	18 (7)	15 (4)	2 (3)
Solid tumors	98 (7)	25 (9)	26 (8)	1 (2)
Lymphoma	45 (3)	10 (4)	9 (3)	5 (8)
Leukemia	27 (2)	11 (4)	1 (0)	1 (2)
ILD	41 (3)	16 (6)	5 (1)	1 (2)
Dialysis	45 (3)	10 (4)	7 (2)	1 (2)
Cirrhosis	19 (1)	4 (1)	7 (2)	0 (0)
hepatic failure	9 (1)	2 (1)	1 (0)	0 (0)
Myeloma	11 (1)	2 (1)	3 (1)	0 (0)
Rheumatic disease	272 (18)	54 (20)	56 (16)	11 (19)
glucocorticoid	364 (24)	79 (29)	87 (26)	27 (46)
Immunologic deficiency	5 (0)	1 (0)	0 (0)	0 (0)

APACHE Acute Physiology and Chronic Health Evaluation

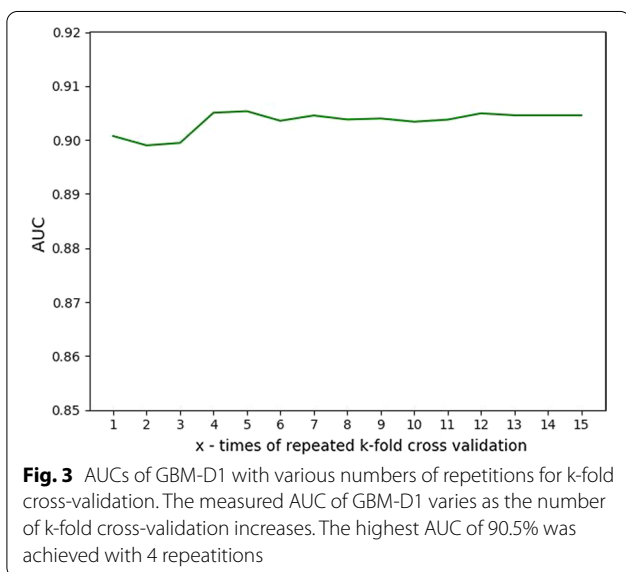
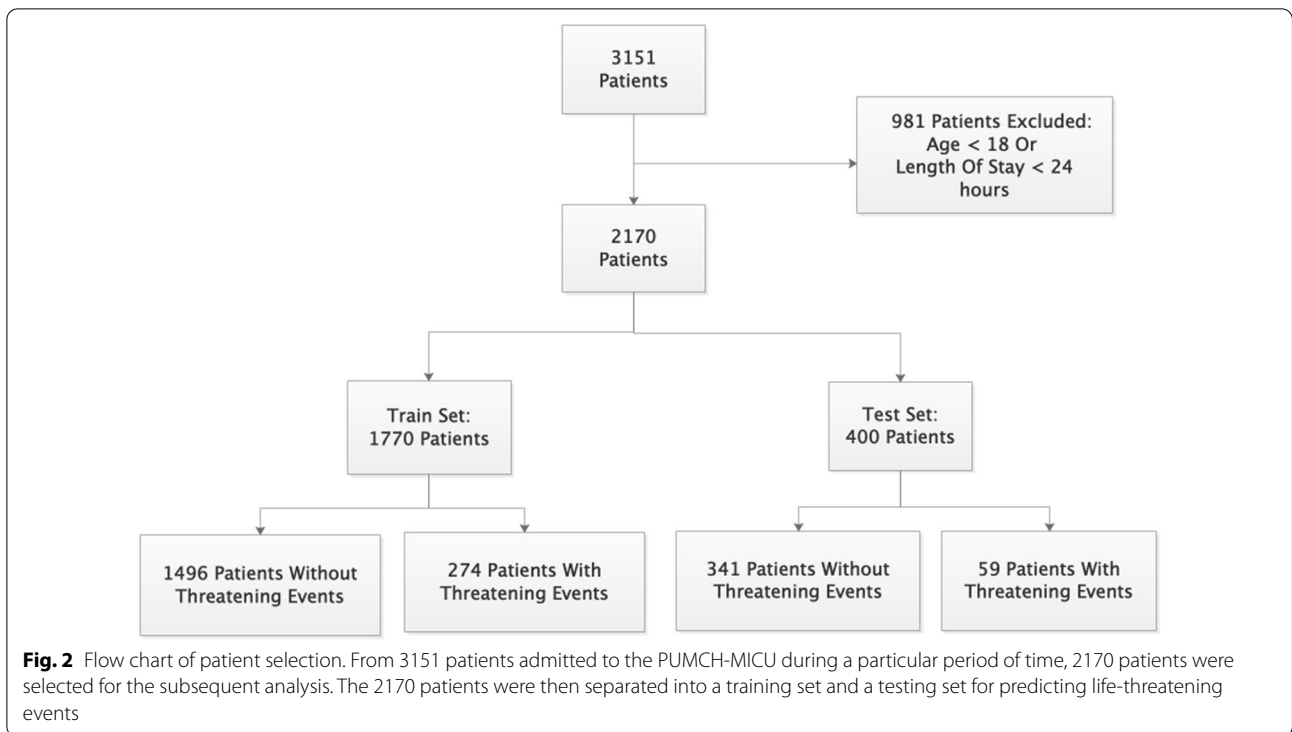
robust as the AUC of GBM-D1 was 90.48% (the AUCs of these predictors see Additional file 1: Fig. S1).

As the different length of stay in ICU may cause the unbalanced number of samples for each patient, the predictors may be biased towards those staying longer in ICU. To evaluate the impact of long ICU stays, a sensitivity analysis was performed on a dataset comprised of one positive sample and one negative sample randomly selected from each patient. The AUC of GBM-D1 was 90.58%, similar to the performance from the

whole dataset (see Additional file 2: Fig. S2 for the plot of AUCs).

Discussion

In this study, the algorithm, lightGBM, was employed to develop our models for predicting life-threatening events due to its superior performance in classification and ability of model interpretation. Single run of k-fold cross-validation might result in noisy estimate of performance, and repeated k-fold cross-validations could obtain



reliable estimates [33]. Thus, to reduce the randomness caused by cross-validation during training, the hyper-parameters were determined by a loop iteration.

Most studies of ICU patients’ prediction model were based on the data in the first 24–72 day of ICU admission to predict the 28-mortality [30] or in-hospital mortality [15, 17]. These prediction models provided few information on daily assessment and treatment. In this study, we

constructed seven predictors for different periods from one day to seven days and included electric defibrillation and chest compression to provide more information in real world. Our predictors enabled dynamic surveillance of ICU patients and identified those at high risk. The metric, AUC, shows the performance decreased with the extended windows of prediction periods. Overall, we could predict life-threatening events for different periods with high accuracy. Given the high complexity of predicting patient outcomes at ICU, we evaluated 4 different prediction methods in the current study. In the actual application of the system, the best predictor will be utilized. Given different feature set and distribution of patient outcomes at different hospitals, we expect no single predictor be the best across all the hospitals and thus evaluation of different ML methods is critical for selecting the best one. GBM-D1 achieved the highest performance (AUC = 0.905 [95% CI 0.863, 0.930]) while GBM-D7 had the lowest performance (AUC = 0.872 [95% CI 0.853, 0.887]). When equal weights were assigned to the sensitivity and specificity, the optimal values can be determined by Youden’s index [34]. Instead of using the default value of 0.5 cutoffs, we applied Youden’s index to determine the cutoff values.

The ranking of features indicates the importance of each feature to prediction. In this study, the rankings of the top features were shown in Fig. 5. For GBM-D1, the top 5 features are norepinephrine, RR_set, Lac, Infusion

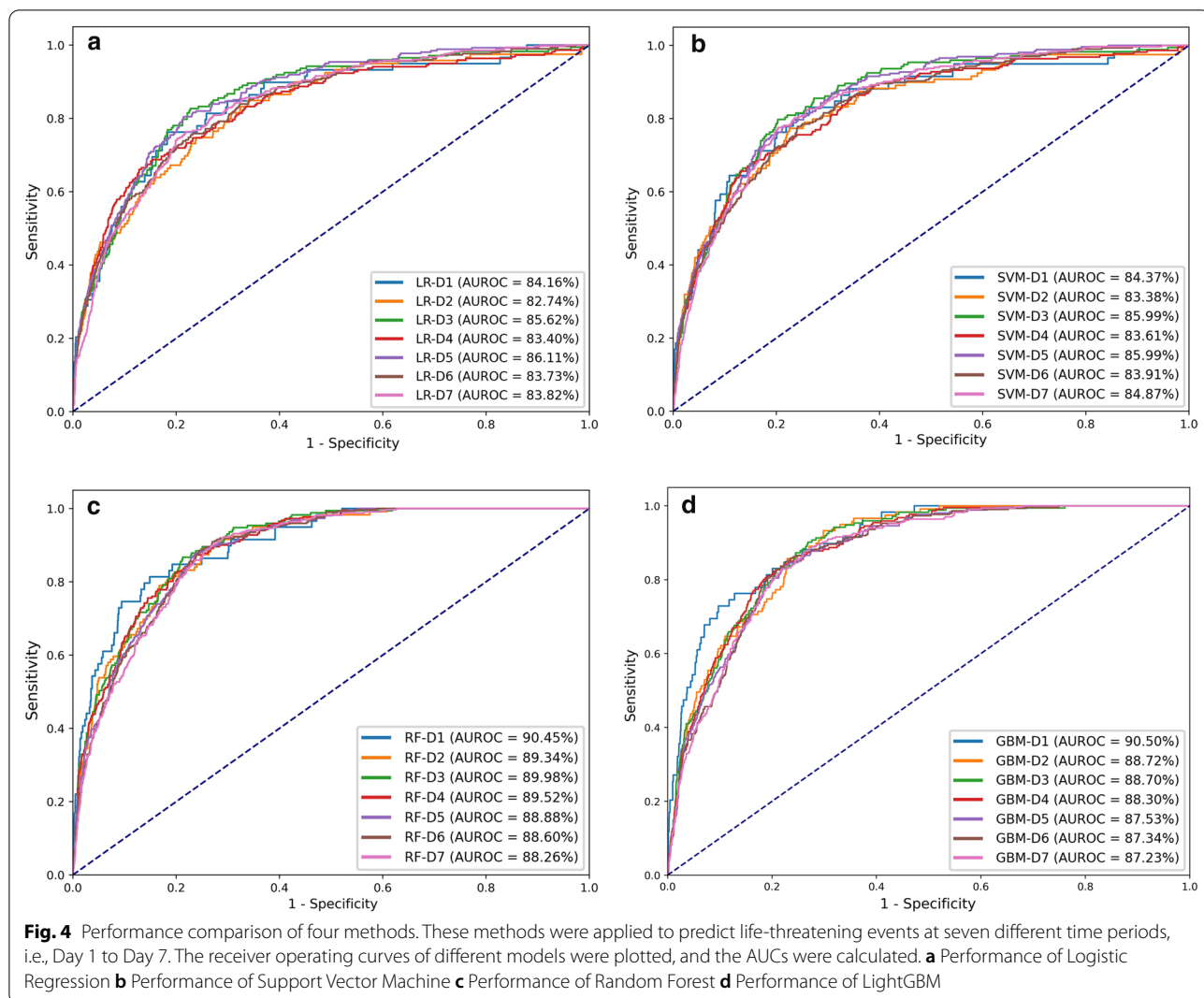


Table 3 Performance comparison for the GBM based predictions. Sensitivity, Specificity and PPV are calculated under a predetermined threshold

	Threshold (Cutoff)(95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	AUC (95% CI)
GBM-D1	0.374 (0.302, 0.437)	0.763 (0.703, 0.906)	0.872 (0.715, 0.915)	0.081(0.042, 0.131)	0.905 (0.863, 0.930)
GBM-D2	0.347(0.338, 0.464)	0.933 (0.785, 0.1)	0.702(0.656, 0.871)	0.087 (0.072, 0.168)	0.887 (0.877, 0.922)
GBM-D3	0.392(0.379, 0.488)	0.902 (0.782, 0.947)	0.731 (0.692, 0.855)	0.130 (0.110, 0.192)	0.887 (0.869, 0.905)
GBM-D4	0.450 (0.427, 0.466)	0.828(0.797, 0.894)	0.793(0.757, 0.821)	0.187 (0.159, 0.216)	0.883 (0.867, 0.899)
GBM-D5	0.455 (0.432, 0.5)	0.839 (0.761, 0.899)	0.776 (0.735, 0.843)	0.205 (0.18, 0.253)	0.875 (0.863, 0.895)
GBM-D6	0.451 (0.429, 0.503)	0.859 (0.77, 0.915)	0.764 (0.705, 0.83)	0.224 (0.182, 0.26)	0.873 (0.855, 0.888)
GBM-D7	0.441 (0.416, 0.474)	0.877 (0.801, 0.907)	0.738 (0.683, 0.784)	0.231 (0.195, 0.263)	0.872 (0.853, 0.887)

CI confidence intervals

pump related fluid input, and RASS. This ranking is not surprising as norepinephrine and infusion pump related fluid input are input events and RR_set is a parameter of the ventilator which is the intervention of treatment. A

high value of lactate indicated the severity of a patient's condition. Age was one of the most important features, and our results showed that age has greater weight in the predictors for a longer prediction window. The ranking

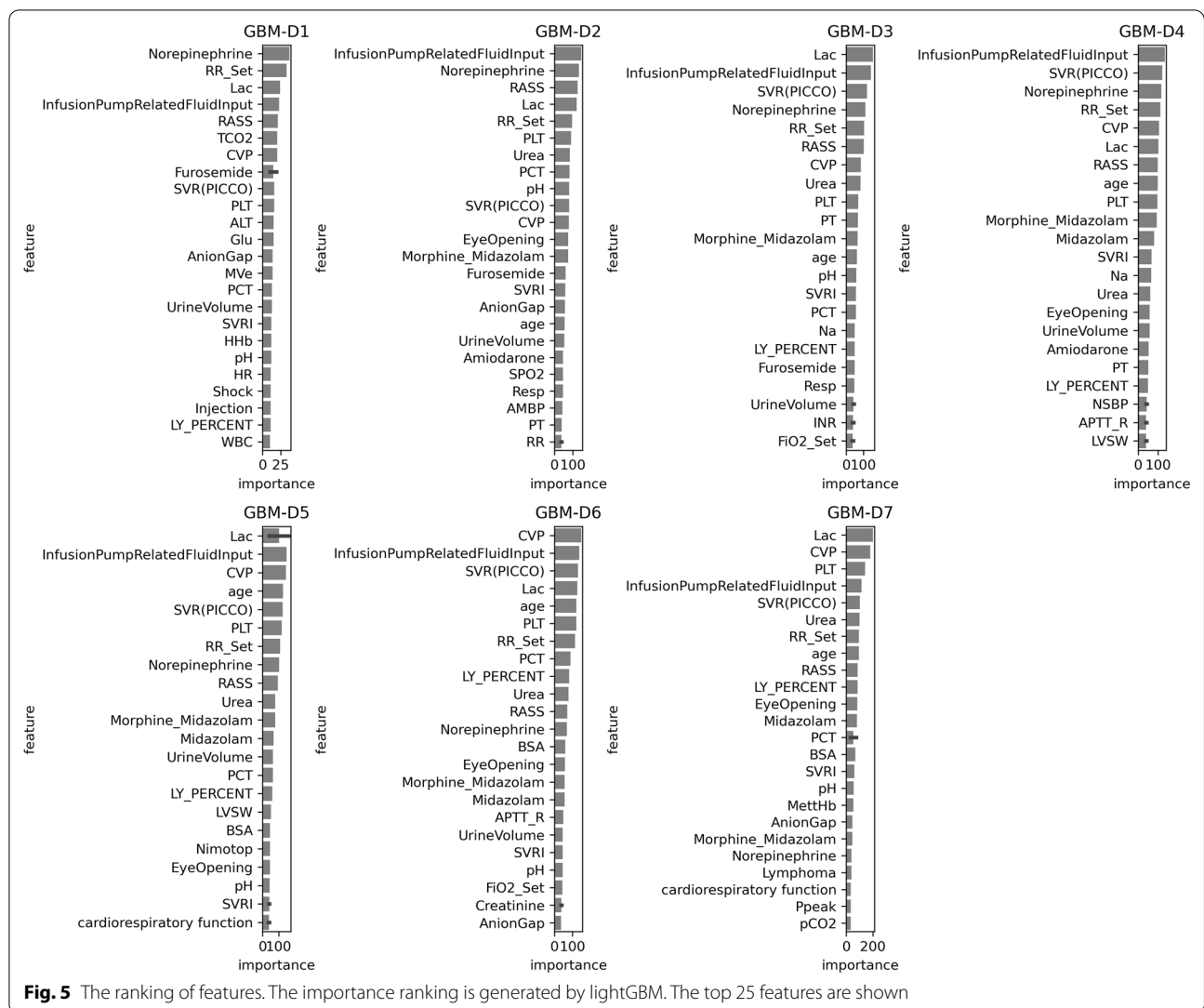


Fig. 5 The ranking of features. The importance ranking is generated by lightGBM. The top 25 features are shown

of other static variables representing the fundamental condition of patients (e.g., cardio-pulmonary function and Lymphoma) increased with extension of prediction window. Therefore, our study revealed some fundamental conditions influencing the outcomes of ICU patients. However, model interpretation was crucial for clinicians to understand the rationale of prediction. The model we developed was used to predict short-term and medium-term outcomes of patients. For the bedside management of hemodynamic unstable patients with norepinephrine, our model might provide more predictive information (e.g., feature importance) about the probability of survival from day 1 to day 7 post norepinephrine initialization. The interpretable predictive model at each window is also crucial for clinicians to utilize the model at bedside. For example, SHapley Additive exPlanations (SHAP) [35] could be used to rank the import features

for each prediction. Clinicians might pay more attention to the highest correlated features which are important for prediction when patients are at high risk. However, it remained a challenge in methodology to unlock black box in machine learning even in some common disease management for intervention, i.e., Hypertension [36].

Our model includes a variety of features beyond the vital sign and represents the real-world condition of patients. To our knowledge, our study is one of the few studies investigating the possibility of predicting both short-term and medium-term life-threatening events for clinicians to decide on treatments and interventions.

There are several limitations in our study. First, the prediction was performed once for all the patients at one time point 6am every day before morning-up. This setup serves as an example to demonstrate the feasibility of the machine learning based prediction of patient outcomes.

However, this model was not built for intervention once any risk factor was identified. Further investigation was urgently needed for any specifically detected high risk factors of an outcome of interest. In the actual clinical application, predictors will be easily retrained for predicting patient outcomes at every hour or at any timepoint (i.e., the real time prediction). In addition, the prediction will be done automatically for a patient whenever a new parameter from him/her comes in. The distribution of outcomes across the different time periods will affect the prediction performance but the impact can be minimized by properly selecting training samples at different classes. In the current study, the distribution of outcome is dynamic with average of approximate 4% of and standard variation of 0.054 for life-threatening events across all time points. Second, our study was based on a single center Medical ICU. Further validation using independent ICU patient cohorts are warranted to generalize the findings. However, the baseline characteristics of the patients in our study suggests that the model developed through this study is highly predictive of the critically-ill patients [16]. Furthermore, Goh et al. showed that the unstructured data can improve the model performance [7], and we will expand our feature set beyond the quantitative features in the future.

Conclusion

In this study, dynamic predictors were developed to predict the risk of life-threatening events in ICU at seven different time points with high accuracy. Such predictions would help clinicians select treatments and interventions in both short term and medium term.

Abbreviations

ALT: Alanine aminotransferase; ADBP: Arterial diastolic blood pressure; AMBP: Arterial mean blood pressure; ARF: Acute renal failure; ASBP: Arterial systolic blood pressure; AUC: Area under curve; BE: Base excess; BIDMC: Beth Israel Deaconess Medical Center; BSA: Body surface area; Ca: Calcium; CI: Cardiac index; Cl: Chloride; COHb: Carboxhemoglobin; COPD: Chronic obstructive pulmonary disease; CPR: Cardiac pulmonary resuscitation; CVP: Central venous pressure; DBil: Bilirubin; EOS_PERCENT: Eosinophil; FIO₂: Fractional inspired oxygen; GBM: Gradient boosting machine; GCS: Glasgow Coma Scale; Glu: Glucose; Hct: Hematocrit; HCT: Hematocrit; HGB: Hemoglobin; HHb: Deoxyhemoglobin; HR: Heart rate; ICU: Intensive care unit; ILD: Interstitial lung disease; INR: International normalized ratio; K: Potassium; Lac: Lactate; LVSW: Left ventricular stroke work; LY_PERCENT: Lymphocyte; MCV: Mean corpuscular volume; MethHb: Methemoglobin; MIMIC: Medical Information Mart for Intensive Care; MONO_ABS: Monocytes absolute; MVe: Minute volume expiration; NA: Sodium; NDBP: Noninvasive diastolic blood pressure; NEUT_PERCENT: Neutrophil; NMBP: Noninvasive mean blood pressure; NSBP: Noninvasive systolic blood pressure; O₂Hb: Oxyhemoglobin; PaO₂: Arterial oxygen pressure; PAP: Pulmonary artery pressure; PCO₂: Pressure of carbon dioxide; pCO₂: Pressure of carbon dioxide; PCT: Procalcitonin; PDW: Platelet distribution width; PH: Potential of hydrogen; PLT: Platelet; pO₂: Pressure of oxygen; Ppeak: Peak pressure; PS: Pressure support; PT: Prothrombin time; PUMCH-MICU: Medical Intensive Care Unit of Peking Union Medical College Hospital; RASS: Richmond Agitation–Sedation Scale; RDW: Red cell distribution width; ROC: Receiver operating curve; RR: Respiratory rate; RR_set: The set of respiratory rate; SI: Stroke index; sO₂: Oxygen saturation; SPO₂: Pulse oximeter oxygen saturation;

SV: Stroke volume; SVR: Systemic vascular resistance; SVRI: Systemic vascular resistance index; TBil: Total bilirubin; TCO₂: Total carbon dioxide; tHb: Total hemoglobin; VTe: Tidal volume expiration; WBC: White blood cell; PICCO: Pulse indicator continuous cardiac output.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-022-02026-x>.

Additional file 1. The AUCs of predictors after excluding the cases of electric defibrillation.

Additional file 2. The AUCs of predictors for randomly selecting one positive sample and one negative sample from each patient.

Acknowledgements

Not applicable

Author contributions

JH and LW designed the study and organized the research. JH, KZH, and FFX performed the data analysis. JH, BD, and LW wrote and edited the manuscript. LW, XHK and JH revised the manuscript. All authors read and approved the final manuscript

Funding

CAMS Innovation Fund for Medical Sciences (CIFMS) 2021-I2M-1-062 from Chinese Academy of Medical Sciences. National Key R&D Program of China 2021YFC2500801 from Ministry of Science and Technology of the People's Republic of China. National Multidisciplinary Cooperative Diagnosis and Treatment Capacity Building Project for Major Diseases.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to the confidentiality policy of Peking Union Medical College Hospital but are available from the corresponding author on reasonable request and with permission of Peking Union Medical College Hospital.

Declarations

Ethics approval and consent to participate

This study was approved by Peking Union Medical College Hospital Institution Review Board (Protocol Number: S-KS44). Informed consent was waived for the retrospective design of the study by the Peking Union Medical College Hospital Institution Review Board. All methods were carried out following relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing Interests

The authors declare no conflict of interest.

Author details

¹Medical Intensive Care Unit, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, 1 Shuai Fu Yuan, Beijing 100730, China. ²Hangzhou Maicim Medical Tech Co., Ltd, Hangzhou, Zhejiang, China.

Received: 18 April 2022 Accepted: 17 October 2022

Published online: 22 October 2022

References

- Bera D, Nayak MM. Mortality risk assessment for ICU patients using logistic regression. *Computing in Cardiology. IEEE*, 2012.
- van Wyk F, Khojandi A, Kamaleswaran R. Improving prediction performance using hierarchical analysis of real-time data: a sepsis case study.

- IEEE J Biomed Health Inform. 2019;23(3):978–86. <https://doi.org/10.1109/JBHI.2019.2894570>.
3. Javan SL, Sepehri MM, Javan ML, Khatibi T. An intelligent warning model for early prediction of cardiac arrest in sepsis patients. *Comput Methods Progr Biomed.* 2019;178:47–58. <https://doi.org/10.1016/j.cmpb.2019.06.010>.
 4. Mollura M, Romano S, Mantoan G, et al. Prediction of Septic Shock Onset in ICU by Instantaneous Monitoring of Vital Signs. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society. IEEE, 2020.
 5. Yuan KC, Tsai LW, Lee KH, et al. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit[J]. *Int J Med Inform.* 2020;141: 104176.
 6. Li X, Xu X, Xie F, et al. A time-phased machine learning model for real-time prediction of sepsis in critical care. *Crit Care Med.* 2020. <https://doi.org/10.1097/CCM.0000000000004494>.
 7. 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.
 8. Zhang, et al. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care.* 2019;23:112. <https://doi.org/10.1186/s13054-019-2411-z>.
 9. Simonov M, Ugwuowo U, et al. A simple real-time model for predicting acute kidney injury in hospitalized patients in the US: a descriptive modeling study. *PLOS Med.* 2019. <https://doi.org/10.1371/journal.pmed.1002861>.
 10. Kim K, Yang H, Yi J, et al. Real-time clinical decision support based on recurrent neural networks for in-hospital acute kidney injury: external validation and model interpretation. *J Med Internet Res.* 2020;23(4):e24120.
 11. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, Chen KL, Yang CY, Lee OK. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care.* 2020;24(1):1–3.
 12. Cherifa M, Blet A, Chambaz A, et al. Prediction of an acute hypotensive episode during an ICU hospitalization with a super learner machine-learning algorithm. *Anesth Analg.* 2020;130(5):1157–66.
 13. Moghadam MC, Masoumi E, Bagherzadeh N, et al. Supervised machine-learning algorithms in real-time prediction of hypotensive events. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) in conjunction with the 43rd Annual Conference of the Canadian Medical and Biological Engineering Society. IEEE, 2020.
 14. Chan B, Chen B, Sedghi A, et al. Generalizable deep temporal models for predicting episodes of sudden hypotension in critically ill patients: a personalized approach. *Sci Rep.* 2020;10(1):1–10.
 15. 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.*
 16. Awad A, Bader-El-Den M, Mcnicholas J, et al. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *Int J Med Inform.* 2017;108:185–95.
 17. Awad A, et al. Predicting hospital mortality for intensive care unit patients: time-series analysis. *Health Inform J.* 2019. <https://doi.org/10.1177/1460458219850323>.
 18. Ray S. A quick review of machine learning algorithms. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). 2019.
 19. Shafaf N, Malek H, et al. Applications of machine learning approaches in emergency medicine; a review article. *Arch Acad Emerg Med.* 2019;7(1):34–34.
 20. 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.
 21. Johnson AEW, Dunkley N, Mayaud L, et al. Patient specific predictions in the intensive care unit using a Bayesian ensemble. IEEE. 2012.
 22. 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.
 23. Lauritsen SM, Kristensen M, Olsen MV, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nat Commun.* 2020;11(1):1–11.
 24. Deshmukh F, Merchant SS. Explainable machine learning model for predicting GI bleed mortality in the intensive care unit. *Am J Gastroenterol.* 2020. <https://doi.org/10.14309/ajg.0000000000000632>.
 25. 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.
 26. Zhang E, Hung SC, Wu CH, Chen LL, Tsai MT, Lee WH. Adverse event and error of unexpected life-threatening events within 24hours of ED admission. *Am J Emerg Med.* 2017; 35(3):479–483
 27. Lin K, Hu Y, Kong G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *Int J Med Inform.* 2019;125:55–61.
 28. Jiang H, Su L, Wang H, Li D, Zhao C, Hong N, Long Y, Zhu W. Noninvasive real-time mortality prediction in intensive care units based on gradient boosting method: model development and validation study. *JMIR Med Inform.* 2021;9(3):e23888. <https://doi.org/10.2196/23888>.
 29. Otaguro T, Tanaka H, Igarashi Y, et al. Machine learning for the prediction of successful extubation among patients with mechanical ventilation in the intensive care unit: a retrospective observational study. *J Nippon Med Sch.* 2021. https://doi.org/10.1272/jnms.JNMS.2021_88-508.
 30. 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.
 31. Qi M. LightGBM: a highly efficient gradient boosting decision tree. In: *Neural Information Processing Systems.* Curran Associates Inc. 2017.
 32. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J.* 2005;47(4):458–72.
 33. Wong TT, Yeh PY. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng.* 2019. <https://doi.org/10.1109/TKDE.2019.2912815>.
 34. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *WILEY-VCH Verl.* 2008;50(3):419–30.
 35. Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: *Nips.* 2017.
 36. Krittana Wong C, Bomback AS, Baber U, et al. Future direction for using artificial intelligence to predict and manage hypertension. *Curr Hypertens Rep.* 2018;20:75.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

