



Machine Learning-Based Model for Predicting Prolonged Mechanical Ventilation in Patients with Congestive Heart Failure

Le Li¹ · Bin Tu¹ · Yulong Xiong¹ · Zhao Hu¹ · Zhenghao Zhang¹ · Shangyu Liu¹ · Yan Yao¹

Accepted: 24 October 2022 / Published online: 16 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Background Mechanical ventilation (MV) is widely used to relieve respiratory failure in patients with congestive heart failure (CHF). Prolonged MV (PMV) is associated with a poor prognosis. We aimed to establish a prediction model based on machine learning (ML) algorithms for the early identification of patients with CHF requiring PMV.

Methods Twelve commonly used ML algorithms were used to build the prediction model. The least absolute shrinkage and selection operator (LASSO) regression was employed to select the key features. We examined the area under the curve (AUC) statistics to evaluate the prediction performance. Data from another database were used to conduct external validation.

Results We screened out 10 key features from the initial 65 variables via LASSO regression to improve the practicability of the model. The CatBoost model showed the best performance for predicting PMV among the 12 commonly used ML algorithms, with favorable discrimination (AUC = 0.790) and calibration (Brier score = 0.154). Moreover, hospital mortality could be accurately predicted using the CatBoost model as well (AUC = 0.844). In the external validation, the CatBoost model also showed satisfactory prediction performance (AUC = 0.780), suggesting certain generalizability of the model. Finally, a nomogram with risk classification of PMV was shown in this study.

Conclusion The present study developed and validated a CatBoost model, which could accurately predict PMV in mechanically ventilated patients with CHF. Moreover, this model has a favorable performance in predicting hospital mortality in these patients.

Keywords Congestive heart failure · prolonged mechanical ventilation · machine learning · prediction model

Introduction

Heart failure (HF) is a global problem with an estimated prevalence of 64.3 million people worldwide, and the most common diagnosis in patients aged 65 years or older admitted to hospital [1, 2]. Although some progress in the treatment, the prognosis of HF is worse than that of most of cancers [3]. Numerous patients with HF are characterized by the presence of tissue congestion due to fluid retention, which is also called congestive heart failure (CHF) [4]. On a pulmonary level, cardiac dysfunction could lead to an increased pulmonary capillary wedge pressure (PCWP), causing interstitial and alveolar edema, with subsequent dyspnea.

Mechanical ventilation (MV) is commonly used in patients with CHF to relieve respiratory failure and reduce work done by the heart [5]. Due to the aging population and the progress in treatments, the MV ratio and the duration of MV have been increasing in recent years [6]. MV is a life-saving treatment; however, it also has some life-threatening

✉ Yan Yao
ianyao@263.net.cn

Le Li
lile_xj@163.com

Bin Tu
312297537@qq.com

Yulong Xiong
xiongyulong@student.pumc.edu.cn

Zhao Hu
huzhao_276@163.com

Zhenghao Zhang
2379508416@qq.com

Shangyu Liu
1811210679@bjmu.edu.cn

¹ Chinese Academy of Medical Sciences, Peking Union Medical College, National Center for Cardiovascular Diseases, Fu Wai Hospital, Beijing 100037, China

complications, including ventilator-associated pneumonia, hemodynamic effects, and sedation agent toxicity [7], and the duration of MV is associated with a near-linear relationship with outcomes [8]. Prolonged MV (PMV) has been defined as 21 days of MV by experts [9]; however, many clinical studies have defined PMV as duration of MV support of >4, >7 or >14 days as these provide greater clinical applicability [10–12]. It is uncontroversial that PMV is strongly associated with poor prognosis [13]. Nevertheless, there is a lack of studies to predict PMV in patients with CHF.

Machine learning (ML), which merges statistical analysis with computer science to produce algorithms, is a subset of artificial intelligence (AI) and could detect relationships between potential explanatory features and a known target outcome [14]. In the previous study, we have developed and validated a CatBoost model, which could accurately predict hospital mortality in mechanically ventilated patients with CHF [15]. In the present study, we aimed to establish a prediction model based on supervised-ML to early identify patients with CHF requiring PMV.

Methods

Sources of Data

Patients' data were collected from the Medical Information Mart for Intensive Care IV (MIMIC-IV, version 1.0), which is a large single-center database containing information of 257,366 individuals who were admitted to various ICUs of the Beth Israel Deaconess Medical Center (BIDMC) between 2008 and 2019 [16]. The code for data extraction is available on GitHub (<https://github.com/MIT-LCP/mimic-iv>). The data of the external validation set were extracted from the eICU Collaborative Research Database (eICU, version 2.0), which is a multi-center ICU database with high granularity data for over 200,000 admissions to ICUs of 208 hospitals located throughout the United States between 2014 and 2015 [17]. The Massachusetts Institute of Technology approved the establishment of the databases with an informed consent exemption. The study was reported according to the recommendations of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [18].

Study Population and Definitions

In the present study, PMV was defined as the duration of MV ≥ 4 days. Specifically, PMV was categorized as prolonged acute MV (PAMV) (duration of MV ≥ 4 days), short-term PMV (duration of MV ≥ 7 days), and long-term PMV (duration of MV ≥ 14 days). The primary purpose was to

build a prediction model for PAMV, and the secondary purposes were to establish the models to predict the short-term PMV and long-term PMV based on the MIMIC-IV cohort.

Participants were selected by the following inclusion criteria: (a) aged > 18 years; (b) with CHF as the major cause for admission; (c) received MV in the first 24-h of ICU admission. Patients with incomplete data were excluded.

Data Collection

PostgreSQL tool (version 13.0) was used to extract medical data from the two databases. The subject IDs were used to identify distinct patients. The demographics, vital signs, common comorbidities, and laboratory tests were included in the initial analysis. In addition, some patients had more than one ICU admission, the data of the first ICU admission and the first medical records during ICU stay were selected for analysis. In the present study, CHF was diagnosed by clinicians based on the guideline of heart failure [19]; pneumonia included community-acquired and hospital-acquired pneumonia, the diagnosis criteria were described elsewhere [20]; acute kidney disease (AKI) was diagnosed based on creatinine or urine output criteria of the clinical guideline [21]; sepsis was diagnosed based on the Sepsis-3 criterion [22]; and ventricular arrhythmia included ventricular tachycardia (sustained or non-sustained), ventricular flutter and ventricular fibrillation. The definitions of other included variables are shown in Supplementary Table S1.

Development of the Prediction Model

Key Features Identification

It is necessary to identify the key features from the training data, which can increase the speed of model fitting and improve the feasibility of the model. In the present study, the least absolute shrinkage and selection operator (LASSO) regression, which could automatically remove unnecessary or unimportant covariates based on a penalization coefficient, was used to eliminate the irrelevant variables [23]. Moreover, the Shapley additive explanations (SHAP) values were used to assess the importance of each feature using a game-theoretic approach using the functions of the SHAP Python package version 0.40.0 [24].

ML Algorithm Selection

A total of 12 widely used ML algorithms including AdaBoost, CatBoost, XGBoost, LightGBM, logistic regression (LR), bootstrap aggregating (Bagging), decision tree, random forest, support vector machine (SVM), naïve Bayes, multi-layer perceptron neural networks (MLP), and K nearest neighbors (KNN) models were used to initially predict

the PMV based on the selected key features. The discrimination of the above models was assessed by the area under the receiver operating characteristic curve (AUC). In addition, accuracy, positive prediction value (PPV), negative prediction value (NPV), balanced accuracy (BA), and F1-score were performed in each model to evaluate the prediction performance. Moreover, the calibration curve and Brier score were applied to qualitatively and quantitatively evaluate the calibration of models, respectively. To demonstrate the decision benefit, the decision curve analysis (DCA) was also conducted. The algorithm with the best discrimination and calibration was selected to perform the remaining analyses.

Hyperparameters Optimization

The hyperparameters, including learning rate and a max of learning depth, are essential for ML algorithms, an ideal combination of hyperparameters could significantly improve the model performance. Accordingly, we performed the hyperparameters optimization (HPO) using the Optuna version 2.10.0, an open-source optimization framework, which enabled us to write complex deep learning experiments quickly, efficiently, and dynamically, to perform the optimization with the Hyperband method and test different combinations of hyperparameters [25]. We performed 100 trials to obtain the best combination of the hyperparameters.

Model Fitting and Evaluation

The final model was fitted with the best combination of the hyperparameters. The data from the MIMIC-IV database were randomly divided into a training set (80%) and an internal validation set (20%). Moreover, the data from the eICU database were used to perform the external validation. To further evaluate the prediction performance of the final model, comparisons of prediction performance between the final model and sequential organ failure assessment (SOFA) score [26], simplified acute physiology score (SAPS-II) [27], logistic organ dysfunction system (LODS) score [28], and a combination of commonly used blood gas indicators, including partial arterial oxygen pressure (PaO₂), partial arterial carbon dioxide pressure (PaCO₂), a saturation of arterial blood oxygen (SO₂), PaO₂, and fraction of inspiration O₂ ratio (PFR), PH, base excess (BE), bicarbonate radical and lactate, were conducted. In addition, PMV is strongly associated with mortality, and prediction of hospital mortality based on the final model was conducted to assess the prediction performance. Finally, a nomogram was performed to visually present the prediction result. Recursive partitioning analysis was performed to construct a survival decision tree for risk stratification with R package ‘rpart’ based on the total points of the nomogram of each participant. Moreover,

the prediction for every patient was plotted in order of their risk to assess the prediction distribution from the model.

Statistical Analyses

The Kolmogorov–Smirnov test was used to evaluate the normal distribution of the data. Continuous variables were expressed as mean \pm standard deviation (SD) and compared using t-test. Levene’s homogeneity of variance test was used to test the assumption of homoscedasticity. If the homoscedasticity was unsatisfied, the Welch’s t-test was used for performing comparisons between groups. Categorical data were expressed as proportions and were compared using the chi-squared test. Additionally, the variance inflation factor (VIF) was assessed among the covariates in the nomogram, and VIF > 4.0 was interpreted as indicating multicollinearity. Variables with VIF > 4.0 were not included in the final model analysis. Univariate and multivariate Logistic regression analyses were performed for all variables of the nomogram to further identify the independent risk factors.

The logic check of the data was conducted to identify the extreme and error value, then to replace them with mean values. Variables with missing values > 30% were excluded. To fill the missing value of the included variables, multiple imputation was performed in the present study [29]. All the statistical analyses were conducted using Python version 3.9.0 (Python Software Foundation, www.python.org) and R software version 4.0.4 (R Foundation for Statistical Computing, Vienna, Austria). A two-tailed test was performed, and a *P*-value < 0.05 was considered to reflect statistical significance.

Results

Baseline Characteristics

A total of 4533 mechanically ventilated patients with CHF were included in the MIMIC-IV database, and all participants were divided into PMV ($n = 1268$) and non-PMV ($n = 3265$) groups based on the duration of MV (Supplementary Fig. S1). In addition, 65 features, including demographics and characteristics, vital signs, therapy, comorbidities, and laboratory tests, were selected for the initial analysis (Supplementary Table S1). After featuring elimination through LASSO regression, 10 key variables, including Glasgow coma scale (GCS) score, PFR, cardiac troponin T (cTnT), N-terminal pro-B type natriuretic peptide (NT-proBNP), diagnosis of pneumonia, sepsis, AKI, anemia, use of inotropic agents, and urine output (UO), were used for establishing the compact model and performing external validation of the eICU database. The variable

comparisons of the non-PMV and the PMV groups of the MIMIC-IV and the eICU databases are shown in Table 1.

Feature Selection

The LASSO regression was performed to identify the key features. The process of feature selection through LASSO regression is shown in Supplementary Fig. S2, with the penalty coefficient (Lambda) increased, the corresponding coefficients of some variables gradually approached zero, which reflected that these variables had little influence on the model. After LASSO regression, 10 key features were selected. The feature importance of the 10 variables was evaluated by SHAP value (Fig. 1). To identify the multicollinearity between the 10 features, the VIF test was performed and showed that there was no significant multicollinearity with all VIFs of the features less than 4.0, the mean VIF was 1.07. Univariate and multivariate Logistic regression analyses were conducted to quantitatively evaluate the influence of the 10 variables on the model. The result suggested that all 10 features had a significant influence on PMV in patients with CHF (Table 2).

Development of the CatBoost Model

Based on the 10 key features, a total of 12 commonly used ML algorithms without parameter optimization were used for the initial analysis, and we found that the CatBoost algorithm had the best prediction model (AUC = 0.790) (Fig. 2). In addition, AUC, accuracy, PPV, NPV, BA, and F1-score of different models based on the 10 features in the internal validation set are summarized in Supplementary Table 3. As shown in the table, the CatBoost model had the best AUC

(0.790), NPV (82.0%), BA (67.9%), and F1 score (0.532). To assess the calibration of the 12 ML models, the calibration curve was conducted as in Fig. 3A, and for simplicity, only LR, XGBoost, and KNN models were selected as the references. We found that the prediction probability of the CatBoost model was the closest to the true probability compared with the others, with a Brier score of 0.154. A satisfactory result was also shown in the DCA (Fig. 3B). Accordingly, the CatBoost model was used to perform the remaining analyses.

HPO based on the 10 features was conducted to improve the prediction performance of the CatBoost model (Supplementary Fig. S3). After 100 trials, the best combination of hyperparameters with the best AUC was obtained. The final settings of the hyperparameter search are listed in Supplementary Table S3. Additionally, comparisons of the full model with the compact model and the final model after HPO were conducted to demonstrate the optimization effect. As shown in Supplementary Fig. S4, the prediction performance of the CatBoost model was significantly improved (AUC: 0.790 vs. 0.817).

Model Evaluation and Validation

To further demonstrate the prediction performance of the CatBoost model, comparisons of the final model with SOFA score, SAPS, blood gas analysis, and LODS score for predicting PAMV in patients with CHF were performed (Fig. 3C). As shown in the figure, the CatBoost model had the best prediction performance compared with others. Moreover, we performed another prediction of short-term MV (7d), long-term MV (14d), and in-hospital death

Table 1 Baseline characteristic of the cohorts

Variables	MIMIC-IV cohort (n = 4533)			eICU cohort (n = 2367)		
	Non-PMV (n = 3265)	PMV (n = 1268)	P value	Non-PMV (n = 903)	PMV (n = 1464)	P value
GCS	13.4 ± 3.8	9.7 ± 4.4	< 0.001	12.7 ± 3.5	11.1 ± 4.0	< 0.001
PFR	204.4 ± 99.5	173.1 ± 91.8	< 0.001	244.8 ± 129.1	220.2 ± 133.3	< 0.001
cTnT, ng/mL	0.99 ± 1.69	1.37 ± 2.14	< 0.001	0.52 ± 1.25	0.96 ± 1.64	< 0.001
NT-proBNP, pg/mL	5804 ± 6007	7954 ± 8198	< 0.001	2451 ± 1943	2686 ± 2189	< 0.001*
Pneumonia, %	1071 (32.8)	753 (59.4)	< 0.001	173 (19.2)	425 (29.0)	< 0.001
Sepsis, %	1215 (37.2)	686 (54.1)	< 0.001	132 (14.6)	388 (26.5)	< 0.001
AKI, %	1747 (53.5)	887 (69.9)	< 0.001	181 (20.0)	428 (29.2)	< 0.001
Anemia, %	2113 (64.7)	917 (72.3)	< 0.001	403 (45.3)	685 (46.8)	0.478
UO, mL/kg/h	0.87 ± 0.91	0.77 ± 0.58	< 0.001	1.41 ± 1.37	1.33 ± 1.36	0.152
Inotropic agent, %	2005 (61.3)	935 (73.7)	< 0.001	247 (27.3)	532 (36.3)	< 0.001

PMV, prolonged mechanical ventilation; GCS, Glasgow coma scale; HR, heart rate; PaO₂, partial arterial pressure of oxygen; PFR, partial arterial pressure of oxygen/ fraction of inspiration O₂; SCr, Serum creatinine; cTnT, cardiac troponin T; NT-proBNP, N-terminal pro-B type natriuretic peptide; AKI, acute kidney injury; VA, ventricular arrhythmia (ventricular tachycardia, ventricular flutter, ventricular fibrillation). (*The eICU cohort did not record NT-proBNP, we used BNP instead)

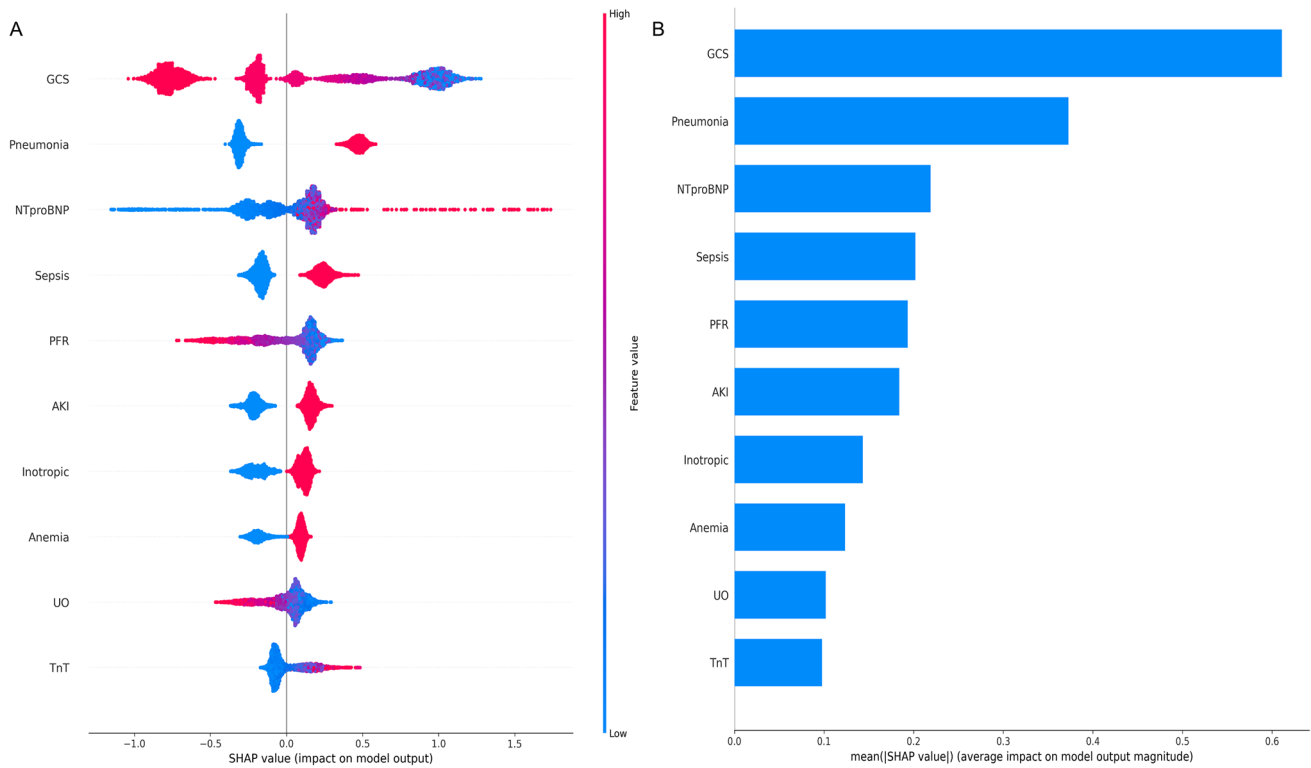


Fig. 1 Features importance assessed by SHAP value. **(A)** The blue to red color represents the feature value (red high, blue low). The x-axis measures the impacts on the model output (right positive, left negative); **(B)** Significance of the predictors in the CatBoost model

Table 2 Univariate and multivariate logistic regression analyses for prolonged acute mechanical ventilation

Variables	Univariate logistic analysis		Multivariate logistic analysis	
	OR (95% CI)	P value	OR (95% CI)	P value
GCS	0.860 (0.847–0.873)	< 0.001	0.879 (0.864–0.893)	< 0.001
PFR	0.996 (0.995–0.997)	< 0.001	0.997 (0.996–0.998)	< 0.001
cTnT	1.104 (1.069–1.141)	< 0.001	1.049 (1.011–1.088)	< 0.001
NT-proBNP	1.430 (1.340–1.501)	< 0.001	1.325 (1.215–1.436)	< 0.001
Pneumonia	2.995 (2.620–3.424)	< 0.001	2.181 (1.876–2.536)	< 0.001
Sepsis	1.989 (1.744–2.268)	< 0.001	1.752 (1.512–2.030)	< 0.001
AKI	2.023 (1.762–2.323)	< 0.001	1.386 (1.184–1.624)	< 0.001
Anemia	1.424 (1.235–1.642)	< 0.001	1.319 (1.124–1.549)	0.001
Urine output	0.771 (0.687–0.864)	< 0.001	0.893 (0.792–0.990)	0.042
Inotropic agents	1.769 (1.533–2.042)	< 0.001	1.492 (1.271–1.753)	< 0.001

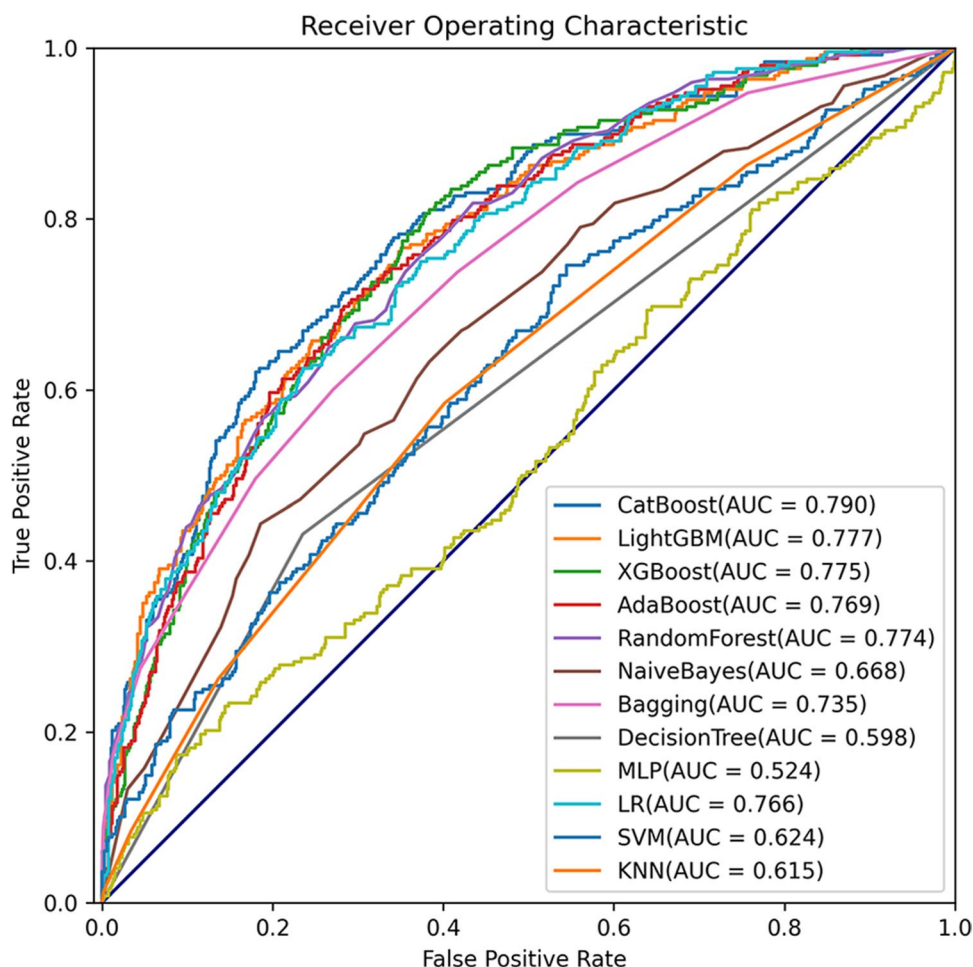
OR, odd ratio; CI, confidence interval; other abbreviations are the same as in Table 1

based on the CatBoost model, and favorable results were obtained in the present study (Fig. 3D-E).

External validation was conducted in this study. As shown in Fig. 3F, in the external validation set with the 10 features, the CatBoost model had the best prediction performance among the 12 algorithms (AUC = 0.780). The accuracy of the CatBoost model in the external validation set was also the highest among the models, with an accuracy of 69.4%.

Finally, a nomogram was performed to visually represent the prediction result. The PAMV probabilities of each patient in the external validation set were obtained using the ‘predict_proba’ function of the CatBoost algorithm. Moreover, the patients were sorted based on the prediction probability. The prediction distribution plot of the CatBoost model with patients sorted in the order of risk showed positive clustering of patients who suffered from PAMV, suggesting the satisfactory discriminatory ability of the model

Fig. 2 Prediction performance of different machine learning algorithms



(Fig. 4A). Furthermore, the decision tree based on total point of each patient in the nomogram was employed to optimize the risk stratification (Fig. 4B). All patients were divided into three groups: low-risk (total points < 186), middle-risk (total points \geq 186 and < 228), and high-risk group (total points \geq 228) (Fig. 4C). In addition, we found that patients in the middle- and high-risk group presented with 3.25 and 9.26-fold risk of PAMV, respectively, with the low-risk group as the reference (Fig. 4D). In addition, the nomograms of prediction of short-term and long-term PMV were shown in Supplementary Figs. S5 and S6.

Discussion

A prediction model of PMV was established based on 12 commonly used ML algorithms. Among the 12 ML algorithms, the CatBoost showed the best prediction performance. In addition, we found that several factors, including pneumonia, GCS score, and NT-proBNP, were strongly associated with PMV in patients with CHF. The prediction performance of the CatBoost model for predicting PMV was

significantly better than the traditional methods, including SOFA score, SAPS, LODS score, and blood gas. Moreover, the CatBoost model showed satisfactory prediction performance both in the internal validation and the external validation sets. Additionally, hospital mortality could be accurately predicted using the final model as well.

The CatBoost algorithm, which belongs to the gradient boosting algorithm family, outperforms other publicly available boosting implementations in terms of quality on a variety of datasets [30]. Two main innovations introduced in CatBoost are the implementation of ordered boosting, a permutation-driven alternative to the classic algorithm, and an innovative algorithm for processing categorical features, both of them can overcome a prediction shift caused by a special kind of target leakage present in all currently existing implementations of gradient boosting algorithms [31]. In addition, CatBoost uses a new schema to calculate leaf values when selecting the tree structure. The schema helps to reduce overfitting, which constrains the generalization ability of machine-learning models [31]. In the present study, the CatBoost model showed satisfactory performance as well.

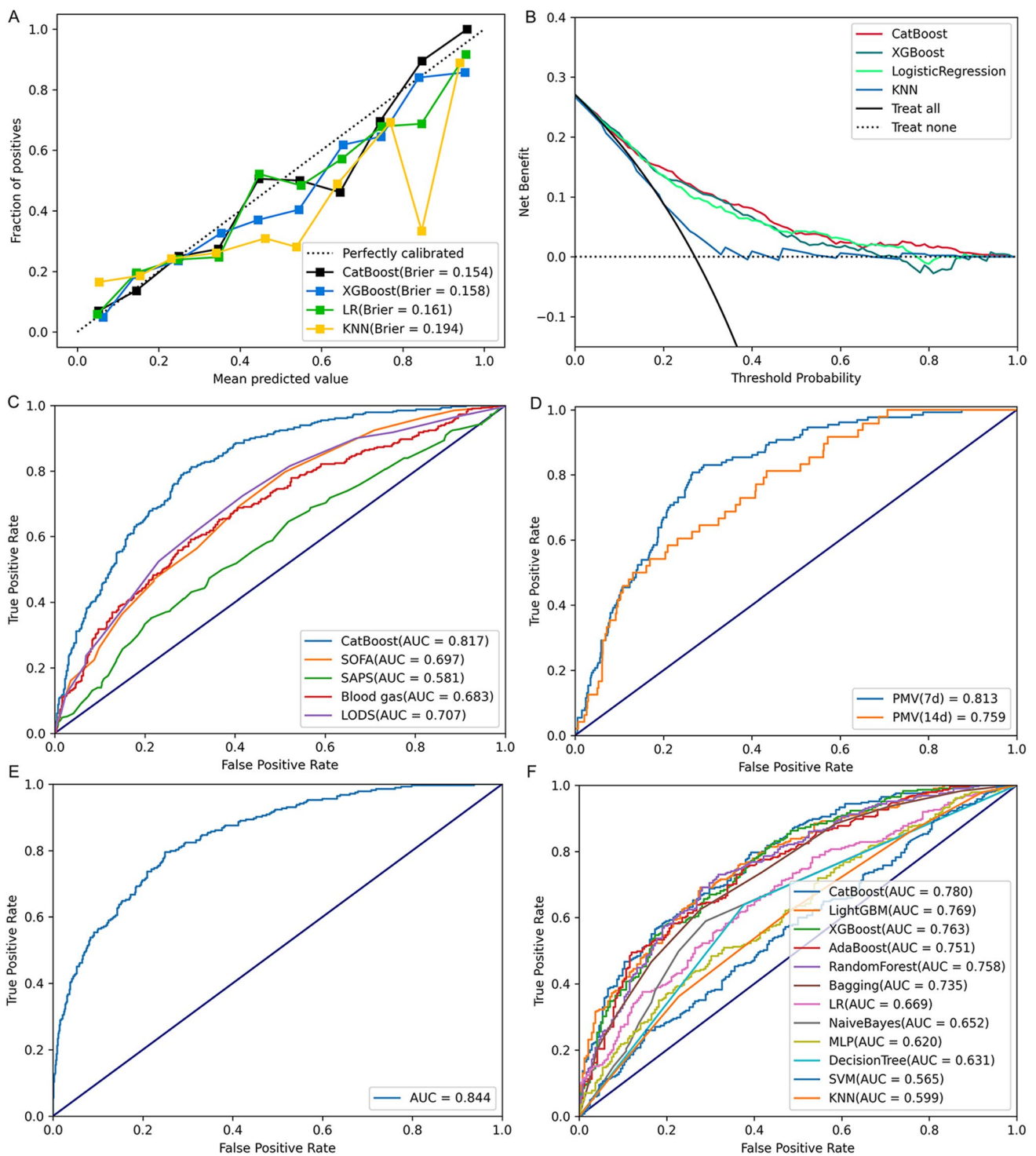


Fig. 3 Model evaluation and validation. (A) Calibration plots; (B) Decision curve analysis; (C) Comparisons of the CatBoost and other models for prolonged acute mechanical ventilation (PAMV) based on the compact 11 features; (D) Prediction of the CatBoost model

for predicting PAMV, short-term prolonged mechanical ventilation (PMV) and long-term PMV; (E) Prediction of in-hospital death based on the CatBoost model; (F) External validation of the prediction model based on the eICU dataset

Stasis of pulmonary vein and capillary is an important pathophysiological feature in patients with CHF, and increased PCWP could lead to interstitial and alveolar edema

and pleural effusion, causing acute respiratory failure [32]. Accordingly, MV is commonly used in patients with CHF to relieve dyspnea. Inflammation of pulmonary parenchyma

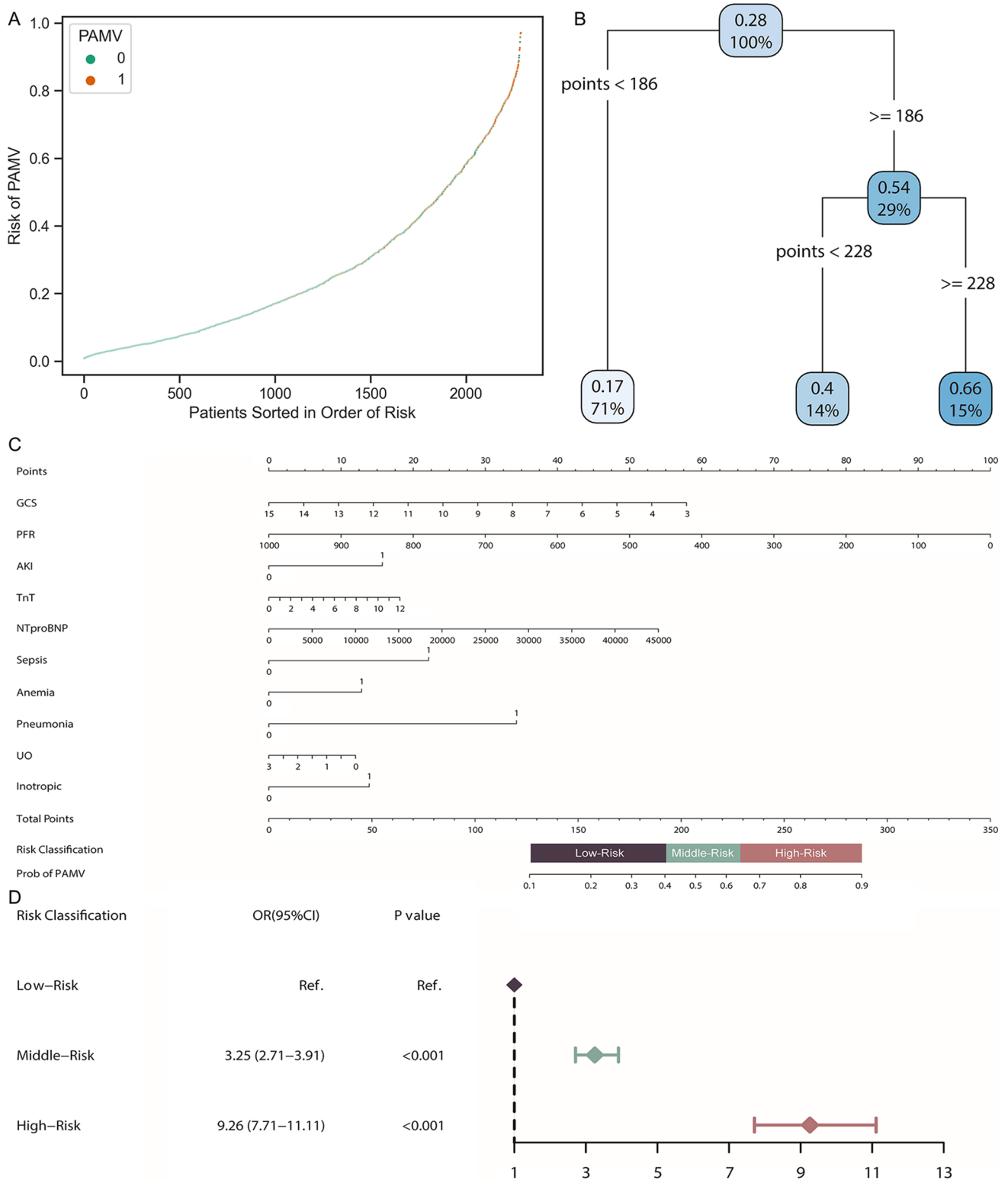


Fig. 4 Model validation and risk stratification. **(A)** Prediction distributions of the risk of PAMV; **(B)** The risk stratification of PAMV according to a decision tree; **(C)** A nomogram of the CatBoost model

for predicting PAMV; **(D)** Logistic analysis of the risk of PAMV based on the risk stratification

and interstitium caused by pneumonia could aggravate lung consolidation and reduce effective respiration area [33], then exacerbate respiratory failure in patients with CHF. In addition, sepsis could aggravate hypoxemia through several mechanisms, including overaction of immune response, circulatory abnormalities, and endothelial dysfunction of the lung [34]. Accordingly, patients with CHF combined with pneumonia or sepsis are likely to suffer from PMV.

NT-proBNP, which is a cleaved proteolytic product, has been shown as an efficient marker in patients with CHF to evaluate prognosis and cardiac dysfunction [35]. Additionally, NT-proBNP could reflect the degree of pulmonary congestion [36]; therefore, NT-proBNP may be associated with respiratory failure in patients with CHF. Moreover, cTnT, which is increased both as a result of ischemic or non-ischemic myocardial injury, is strongly associated with cardiac dysfunction [37]. CHF reduces renal perfusion and then causes AKI, and renal dysfunction, in turn, increases the absorption of sodium water, placing a heavy burden on the heart. Therefore, the renal injury will deteriorate cardiac dysfunction in patients with CHF. Our findings suggested that NT-proBNP, cTnT, UO, and AKI were strongly associated with PMV as expected. In addition, use of inotropic agents and GCS scores were also found as the prediction factors for PMV in this study, and the results would be helpful in future research.

Blood gas analysis, SOFA score, LODS score, and SAPS-II are widely used to evaluate the respiratory status and the severity of diseases. We conducted comparisons of the CatBoost model with these models to predict PMV and found that the CatBoost model was significantly superior to these models. The previous study has demonstrated that PMV was significantly associated with mortality [13]. Accordingly, we hypothesize that hospital mortality could be accurately predicted using the final model which is used to predict PMV. The results showed that the CatBoost model exhibited favorable prediction performance to predict in-hospital mortality based on the data of the selected 10 features (AUC = 0.844). In addition, external validation was performed in the present study. We found that the prediction performance of the CatBoost model was also excellent in the eICU database.

The decision tree algorithm based on the nomogram was used to optimize the risk stratification of PMV in CHF patients (Fig. 4C). Due to the 10 features for prediction were readily accessible and frequently monitored in routine clinical practice, CHF patients with high-risk of PMV could be easily identified using this nomogram. Our prediction model may lead to several changes in the management of mechanically ventilated patients with CHF such as prediction of ICU workload and subsequent cost incurred by the hospital,

and the ability to rationalize and tailor mechanical ventilation strategy for high-risk patients. Moreover, the CatBoost model may facilitate decision making regard to the timing of tracheostomy in the high-risk of PMV patients. Timely tracheostomy not only had shorter duration of MV and intensive care unit length of stay but also markedly lower hospital mortality [38]. Accordingly, we suggest that our model certainly had clinical application value in the early identification of PMV in patients with CHF and in the administration of appropriate preventive treatments.

Limitations

Several limitations should be considered in this study. First, the data for analysis were collected from two public databases. There were some important variables, including left ventricular ejection fraction, with a high rate of missing values were excluded, which may affect the prediction performance. Second, there is no consensus on the accepted definition of PMV in CHF patients, while other investigators have used a wide range of time frames within a range of 4 to 21 days to define PMV. Considering the progresses in the management of CHF, the decompensated symptoms of CHF could be more rapidly corrected than before, we defined PMV as MV \geq 4 days in this study. Third, this was a retrospective cohort study, and further prospective studies are needed to confirm the findings. Accordingly, the results should be interpreted with caution.

Conclusion

A prediction model of PMV with satisfactory prediction performance was established based on 12 commonly used ML algorithms. The CatBoost model was better than other predictive methods in predicting PMV in patients with CHF. Additionally, the CatBoost model may help in predicting in-hospital mortality in these patients. These findings need to be confirmed in future prospective studies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10557-022-07399-9>.

Authors' Contributions This study was designed by LL. ZZH, YLX, ZH, SYL, and BT were responsible for data collation and statistical analysis. LL wrote the first draft. YY reviewed and checked the manuscript. All authors read and approved the final manuscript.

Funding This project was supported by the High-level Hospital Project of Fuwai Hospital, Chinese Academy of Medical Sciences (2022-GSP-GG-25).

Data availability The data in of this study could be found in <https://physionet.org/content/mimiciv/1.0/> and <https://physionet.org/content/eicu-crd/2.0/>

Code Availability The code can be available upon reasonable request to the corresponding author.

Declarations

Ethics Approval and Consent to Participate Because the study was an analysis of the third-party anonymized publicly available database with pre-existing institutional review board (IRB) approval, IRB approval from our institution was exempted.

Consent for Publication All the authors have agreed to publish this article.

Conflict of Interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Diseases GBD, Injuries C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396:1204–22.
2. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. *Eur J Heart Fail*. 2020;22:1342–56.
3. Braunwald E. The war against heart failure: the Lancet lecture. *Lancet*. 2015;385:812–24.
4. La Franca E, Manno G, Ajello L, et al. Physiopathology and diagnosis of congestive heart failure: consolidated certainties and new perspectives. *Curr Probl Cardiol*. 2021;46:100691.
5. Alviar CL, Miller PE, McAreavey D, et al. Positive pressure ventilation in the cardiac intensive care unit. *J Am Coll Cardiol*. 2018;72:1532–53.
6. Herring AA, Ginde AA, Fahimi J, et al. Increasing critical care admissions from U.S. emergency departments, 2001-2009. *Crit Care Med*. 2013;41:1197–204.
7. Pham T, Brochard LJ, Slutsky AS. Mechanical ventilation: state of the art. *Mayo Clin Proc*. 2017;92:1382–400.
8. MacIntyre NR, Epstein SK, Carson S, et al. Management of patients requiring prolonged mechanical ventilation: report of a NAMDRC consensus conference. *Chest*. 2005;128:3937–54.
9. Lamas D. Chronic critical illness. *N Engl J Med*. 2014;370:175–7.
10. Zilberberg MD, Luippold RS, Sulsky S, Shorr AF. Prolonged acute mechanical ventilation, hospital resource utilization, and mortality in the United States. *Crit Care Med*. 2008;36:724–30.
11. Seneff MG, Zimmerman JE, Knaus WA, Wagner DP, Draper EA. Predicting the duration of mechanical ventilation. The importance of disease and patient characteristics. *Chest*. 1996;110:469–79.
12. Lin WC, Chen CW, Wang JD, Tsai LM. Is tracheostomy a better choice than translaryngeal intubation for critically ill patients requiring mechanical ventilation for more than 14 days? A comparison of short-term outcomes. *BMC Anesthesiol*. 2015;15:181.
13. Damuth E, Mitchell JA, Bartock JL, Roberts BW, Trzeciak S. Long-term survival of critically ill patients treated with prolonged mechanical ventilation: a systematic review and meta-analysis. *Lancet Respir Med*. 2015;3:544–53.
14. Bzdok D, Krzywinski M, Altman N. Machine learning: supervised methods. *Nat Methods*. 2018;15:5–6.
15. Li L, Zhang Z, Xiong Y, et al. Prediction of hospital mortality in mechanically ventilated patients with congestive heart failure using machine learning approaches. *Int J Cardiol*. 2022;358:59–64.
16. Johnson ABL, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV (version 1.0), PhysioNet, 2021.
17. Pollard TJ, Johnson AEW, Raffa JD, et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data*. 2018;5:180178.
18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
19. Hunt SA, Baker DW, Chin MH, et al. ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: executive summary a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure): developed in collaboration with the International Society for Heart and Lung Transplantation; Endorsed by the Heart Failure Society of America. *Circulation*. 2001;104:2996–3007.
20. Lanks CW, Musani AI, Hsia DW. Community-acquired pneumonia and hospital-acquired pneumonia. *Med Clin North Am*. 2019;103:487–501.
21. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract*. 2012;120:c179–84.
22. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. 2016;315:801–10.
23. Kok BC, Choi JS, Oh H, Choi JY. Sparse extended redundancy analysis: variable selection via the exclusive LASSO. *Multivar Behav Res*. 2021;56:426–46.
24. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56–67.
25. Li L, Jamieson K, Desalvo G, Rostamizadeh A, Talwalkar A. Hyperband: a novel bandit-based approach to hyperparameter optimization. *J Mach Learn Res*. 2016;18:1-52.
26. Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA*. 2001;286:1754–8.
27. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1993;270:2957–63.
28. Le Gall JR, Klar J, Lemeshow S, et al. The logistic organ dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. *JAMA*. 1996;276:802–10.
29. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med*. 2016;4:30.
30. Hancock JT, Khoshgoufar TM. CatBoost for big data: an interdisciplinary review. *J Big Data*. 2020;7:94.
31. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett

- R (Ed.), *Advances in neural information processing systems* 31, Curran Associates, Inc., 2018;6638–6648.
32. Boorsma EM, Ter Maaten JM, Damman K, et al. Congestion in heart failure: a contemporary look at physiology, diagnosis and treatment. *Nat Rev Cardiol.* 2020;17:641–55.
 33. Alcon A, Fabregas N, Torres A. Pathophysiology of pneumonia. *Clin Chest Med.* 2005;26:39–46.
 34. Lelubre C, Vincent JL. Mechanisms and treatment of organ failure in sepsis. *Nat Rev Nephrol.* 2018;14:417–27.
 35. Goetze JP, Bruneau BG, Ramos HR, et al. Cardiac natriuretic peptides. *Nat Rev Cardiol.* 2020;17:698–717.
 36. Omland T. Advances in congestive heart failure management in the intensive care unit: B-type natriuretic peptides in evaluation of acute heart failure. *Crit Care Med.* 2008;36:S17–27.
 37. Park KC, Gaze DC, Collinson PO, Marber MS. Cardiac troponins: from myocardial infarction to chronic disease. *Cardiovasc Res.* 2017;113:1708–18.
 38. Rumbak MJ, Newton M, Truncale T, et al. A prospective, randomized, study comparing early percutaneous dilational tracheotomy to prolonged translaryngeal intubation (delayed tracheotomy) in critically ill medical patients. *Crit Care Med.* 2004;32:1689–94.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.