

Use of machine learning algorithms to predict life-threatening ventricular arrhythmia in sepsis

Le Li, Zhuxin Zhang, Likun Zhou, Zhenhao Zhang, Yulong Xiong, Zhao Hu, and Yan Yao *

Chinese Academy of Medical Sciences, Peking Union Medical College, National Center for Cardiovascular Diseases, Beilishi Road 167#, Xicheng District, Beijing, China

Received 16 January 2023; revised 24 March 2023; accepted 5 April 2023; online publish-ahead-of-print 6 April 2023

Aims

Life-threatening ventricular arrhythmias (LTVAs) are common manifestations of sepsis. The majority of sepsis patients with LTVA are unresponsive to initial standard treatment and thus have a poor prognosis. There are very limited studies focusing on the early identification of patients at high risk of LTVA in sepsis to perform optimal preventive treatment interventions. We aimed to develop a prediction model to predict LTVA in sepsis using machine learning (ML) approaches.

Methods and results

Six ML algorithms including CatBoost, LightGBM, and XGBoost were employed to perform the model fitting. The least absolute shrinkage and selection operator (LASSO) regression was used to identify key features. Methods of model evaluation involved in this study included area under the receiver operating characteristic curve (AUROC), for model discrimination, calibration curve, and Brier score, for model calibration. Finally, we validated the prediction model both internally and externally. A total of 27 139 patients with sepsis were identified in this study, 1136 (4.2%) suffered from LTVA during hospitalization. We screened out 10 key features from the initial 54 variables via LASSO regression to improve the practicability of the model. CatBoost showed the best prediction performance among the six ML algorithms, with excellent discrimination (AUROC = 0.874) and calibration (Brier score = 0.157). The remarkable performance of the model was presented in the external validation cohort ($n = 9492$), with an AUROC of 0.836, suggesting certain generalizability of the model. Finally, a nomogram with risk classification of LTVA was shown in this study.

Conclusion

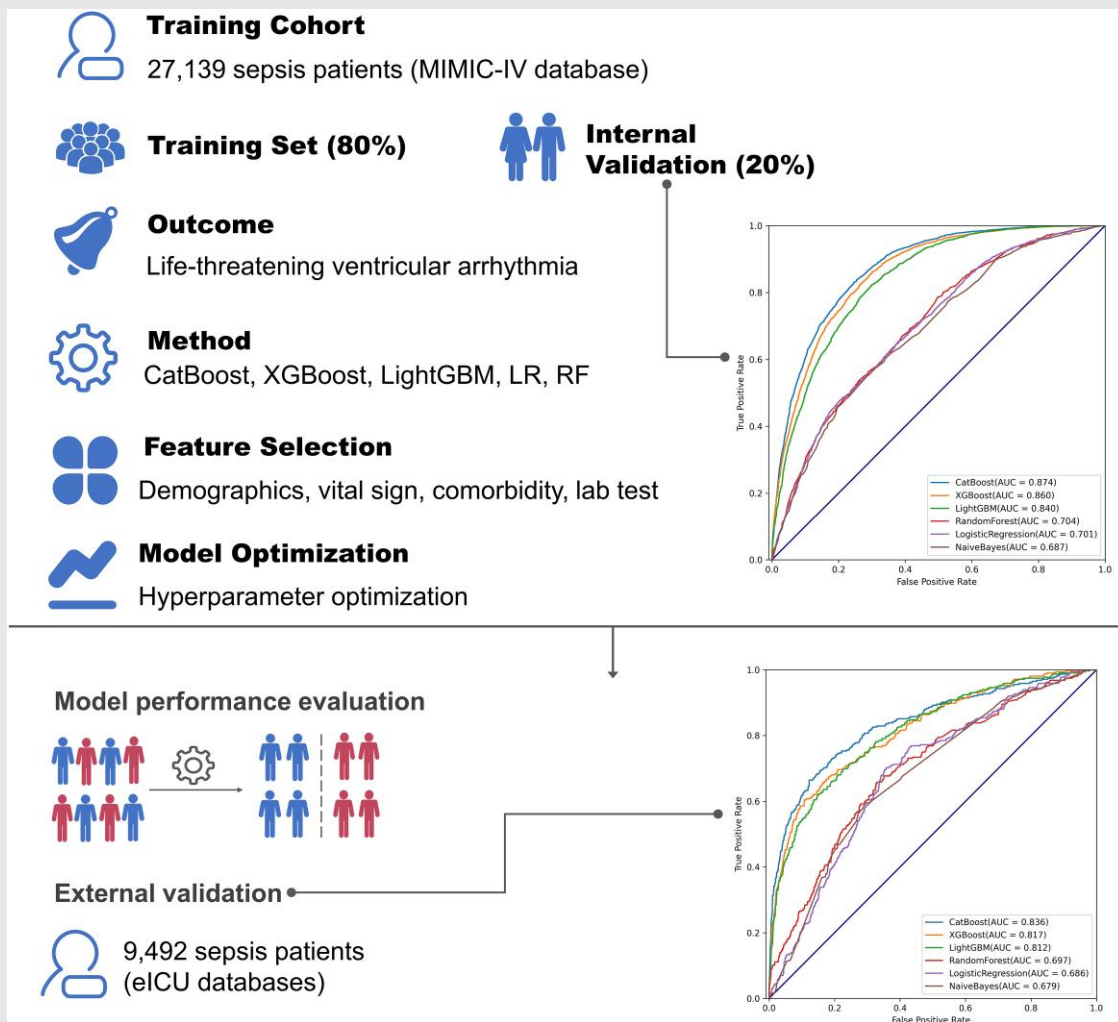
We established and validated a machine learning-based prediction model, which was conducive to early identification of high-risk LTVA patients in sepsis, thus appropriate methods could be conducted to improve outcomes.

* Corresponding author. Tel: +86 010-88398591, Fax: +86 010-88322401, Email: ianyao@263.net.cn

© The Author(s) 2023. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Graphical Abstract



Overview of the study design. A total of 27 139 patients with sepsis were randomly divided into the training set and internal validation set at a ratio of 8:2. Several machine learning algorithms including CatBoost, XGBoost, and logistic regression were used to perform the model fitting. After key feature selection, hyper-parameter optimization was implemented to modify the model. Finally, 9492 sepsis patients from another database were involved to conduct external validation.

Keywords

Sepsis • Life-threatening ventricular arrhythmia • Risk stratification • Machine learning • Prediction model

Introduction

Sepsis, defined as an infection associated with organ injury distant from the site of infection, is characterized as fever (or hypothermia), tachycardia, tachypnoea, and blood leucocyte changes.¹ Moreover, sepsis is a common condition in the intensive care unit (ICU) and is associated with unacceptably high mortality, approximately 2.8 million deaths per year are attributable to sepsis.² Robust data indicate that patients with sepsis are vulnerable to suffering from ventricular arrhythmias (VAs), including ventricular tachycardia (VT) and ventricular fibrillation (VF), which could explain the majority of sudden cardiac death (SCD) cases.^{3,4} Furthermore, existing evidence has confirmed that patients hospitalized for sepsis developed new VA events and were at high risk of SCD and in-hospital mortality.^{5,6}

Life-threatening ventricular arrhythmias (LTVAs), which occur with severely depressed ventricular function and unstable haemodynamics

state, are the main causes of cardiac arrest.^{7,8} However, half of the patients with LTVA are unresponsive to initial standard treatment, and thus have a poor prognosis.⁹ Moreover, the patients requiring more than 40 min of cardiopulmonary resuscitation almost all died.¹⁰ Therefore, early recognition of sepsis patients at high risk of LTVA is critical for timely interventions as well as intensive care and monitoring.

Machine learning (ML), a sub-field of artificial intelligence, can easily incorporate numerous variables and then identify non-linear associations and complex interactions among variables.¹¹ ML algorithms have been explored in extensive fields of arrhythmia, including, but not limited to, classification of arrhythmia, detection of arrhythmia origins, severity risk prediction, and prognosis prediction in certain populations.^{12–14} Nevertheless, the utility of ML algorithms in identifying sepsis patients with a high risk of LTVA has not been determined.

Given the susceptibility and the severity of LTVA in sepsis patients, we aimed to establish a novel model to predict the incidence of

LTVA in patients with sepsis utilizing ML algorithms. The validated capacity of enabling straightforward and accurate risk stratification of LTVA in sepsis patients of the prediction model may facilitate to perform optimal preventive treatment interventions.

Methods

This was a retrospective study based on two large-scale databases, aiming to build a clinical prediction model to predict LTVA in sepsis patients (see [Graphical Abstract](#)).

Sources of data

A large critical care database, named Medical Information Mart for Intensive Care IV (MIMIC-IV, version 2.0), was employed in the present study to perform model training. The MIMIC-IV database contains data from more than 200 000 individuals who were admitted to various ICUs of the Beth Israel Deaconess Medical Center (BIDMC) between 2008 and 2019.¹⁵ Additionally, data of sepsis patients extracted from the eICU Collaborative Research Database (eICU, version 2.0), which was a multi-centre critical care database with high granularity data for over 200 000 admissions to ICUs, were used to conduct external validation. Data of the two databases include vital signs, medications, laboratory measurements, observations and notes charted by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, survival data, and more. Data from MIMIC-IV and eICU were independent of each other. The description of eICU was available elsewhere.¹⁶ The author (L.L.) is qualified to inquiring the information from the databases (record ID: 35965741). Since 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. The study was reported according to the recommendations of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement.¹⁷

Participants

We included 27 139 and 9492 sepsis patients with known outcomes (LTVA or non-LTVA) in training and validation cohorts, respectively. Patients aged < 18 years old or with inadequate data were excluded. In the present study, sepsis was defined as infection-related organ dysfunction which was represented by an increase in the Sequential Organ Failure Assessment (SOFA) score of 2 points or more.¹⁸ Furthermore, in this study, LTVA came in two main forms: sustained VT (duration greater than 30 s) and VF.¹⁹

Data imputation

The extreme and error data that failed the logic check were filled with mean values. We excluded features with a ratio of missing value > 30%. Mean imputation was performed to fill in missing data of <5%. Multiple imputation was used to handle variables with missing data of between 5 and 30%.²⁰

Data collection

A total of 67 features including demographic, clinical, and laboratory data from the two databases were extracted via the PostgreSQL tool (version 13.0) (see [Supplementary material online, Table S1](#)). To ensure the practicability of the prediction model, we employed the data of the first medical records instead of the maximum or the minimum data during ICU stay.

Model development

Feature selection

After excluding the features with a heavy proportion of missing value and imputing missing data, 54 features were remaining for feature selection. To eliminate irrelevant and redundant collinear features and improve the practical applicability of the model, we conducted key feature selection employing least absolute shrinkage and selection operator (LASSO) regression which could automatically remove unnecessary or uninformative covariates based on a penalization coefficient.²¹ Additionally, the SHapley Additive exPlanations (SHAP) values²² and random forest (RF) algorithm²³ were applied to rank all features by feature importance. Finally, the most predictive variables were selected by integrating the three methods.

Model evaluation

To evaluate the performance of ML models, we used two metrics: area under the receiver operating characteristic curve (AUROC) and area under the precision–recall curve (AUPR). The calibration curve and Brier score were applied to qualitatively and quantitatively evaluate the calibration of models, respectively. Decision curve analysis (DCA) was performed to present the decision benefit of models. Additionally, accuracy, sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), Matthews correlation coefficient (MCC), and F1-score were performed in each model to quantitatively evaluate prediction performance.

ML algorithm selection

A total of six commonly used baseline ML algorithms, including CatBoost, XGBoost, LightGBM, logistic regression (LR), RF, and naïve Bayes, were employed to initially predict LTVA in sepsis patients based on the selected features. The algorithm with the best prediction performance was selected for the remaining analyses.

Model optimization

In ML algorithms, hyper-parameter optimization (HPO) is a critical procedure for achieving satisfactory performance. Hyper-parameters are numerical pre-sets whose values are assigned before the commencement of a learning process. An ideal combination of hyper-parameters will significantly improve model performance. To visualize the process of HPO, the Optuna package (version 2.10.0),²⁴ an open-source optimization framework, which allowed us to write complex ML experiments quickly, efficiently, and dynamically, was employed to perform HPO with the Hyperband method and test different combinations of hyper-parameters.²⁵

Model fitting

Before model fitting, the combination of Synthetic Minority Over-sampling Technique and under-sampling was used to overcome the class-imbalance issue in the present study.²⁶ The final model was fitted with the best combination of the hyper-parameters. Data from the training cohort were randomly divided into the training set and the internal validation set based on the ratio of 8:2. Herein, we modelled the LTVA prediction task as a binary classification problem. All included ML models output a normalized probability of LTVA risk ranging from 0 to 1.

Model validation

External validation based on the eICU cohort was performed to assess the robustness and generalizability of the model. AUROC was used to evaluate the performance of the six ML algorithms. Because LTVA was highly associated with an increased risk of mortality, the CatBoost model was employed to predict 1-year mortality. Moreover, the prediction results were validated and presented with a nomogram based on the external validation set. 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. Additionally, the prediction for every patient was plotted in order of their risk to assess the prediction distribution from the model.

Statistical analysis

Statistical analysis was performed in Python (version 3.9.0) and R (version 4.1.0). Patients were divided into LTVA and non-LTVA groups, and variables were compared between groups. Descriptive statistics were reported as frequency count and percentage, for qualitative data, and mean \pm standard deviation, for quantitative data. t-Test and chi-square test were used to evaluate differences between two groups for continuous and categorical variables, respectively. Spearman correlation coefficient was employed to verify the possible statistical association between variables. *P* values ≤ 0.05 were considered significant.

Results

Characteristics of the study population

A total of 27 139 patients [mean (SD) age, 68.2 (16.2) years; 15 626 (57.6%) male] with sepsis were identified in the training cohort, among them 1136 patients (4.2%) suffered from LTVA during hospitalization

(see [Supplementary material online, Figure S1](#)). For patients with LTVA, mean temperature (36.7 ± 0.7 vs. 36.9 ± 0.6 , $P < 0.001$) and systolic blood pressure (SBP) (85.0 ± 17.3 vs. 89.5 ± 16.8 , $P < 0.001$) were significantly higher than patients without LTVA. Additionally, in the LTVA group, 734 patients (63.7%) presented with congestive heart failure (CHF), 579 (51.0%) with atrial fibrillation (AF), 390 (34.3%) with acute myocardial infarction (AMI), and 533 (46.9%) with pneumonia, incidences of the comorbidities were statistically higher than the non-LTVA group. The baseline characteristics of the validation cohort are summarized in [Table 1](#). The comparisons of baseline characteristics between the MIMIC-IV cohort and the eICU cohort are shown in [Supplementary material online, Table S2](#).

Development of the prediction model

The LASSO regression was employed to eliminate redundant or irrelevant features, and to improve the practicability of the model. Only 10 of the 54 features were eventually chosen for modelling, among which 9 features had a positive association with LTVA [SBP, Charlson comorbidity index (CCI), requirement of mechanical ventilation (MV), quick SOFA (qSOFA) score, pneumonia, CHF, AF, AMI, acute kidney injury (AKI)] ([Figure 1](#)). Furthermore, feature importance was evaluated by SHAP value and RF algorithm and showed the analogous results of LASSO regression (see [Supplementary material online, Figure S2](#)). In addition, SHAP force plot enabled personalized interpretation of the model (see [Supplementary material online, Figure S3](#)). Accordingly, the 10 key features were selected to build the prediction model. The correlation between the 10 features was assessed by Spearman's correlation coefficient (see [Supplementary material online, Figure S4](#)).

We compared the initial prediction performance (without model optimization) of the 6 ML algorithms based on the 10 features and found that the CatBoost algorithm presented with the best performance with an AUROC of 0.874 ([Figure 2](#)). [Table 2](#) summarizes commonly used evaluation indicators of different ML algorithms. The CatBoost presented with the best AUROC (0.874), accuracy (0.794), sensitivity (0.710), specificity (0.849), PPV (0.755), NPV (0.818), MCC (0.566), and F1-score (0.732). The precision–recall curve showed the trade-off between precision and recall, which, in this case, means the fraction of patients who developed LTVA in the patients who were predicted to have LTVA (precision) and their fraction in all the patients who developed LTVA (recall). The comparisons of AUPR between CatBoost and other ML algorithms were conducted in the present study. The result

showed that the CatBoost algorithm had the best AUPR (0.842) (see [Supplementary material online, Figure S5](#)).

Calibration refers to the degree of agreement between the predicted probabilities and the actual probabilities, which can be evaluated both qualitatively and quantitatively. Qualitative evaluation is typically performed using calibration curves, while quantitative evaluation is often done using the Brier score. Brier score is calculated based on the Euclidean distance between the actual outcome and the predicted probability assigned to the outcome for each observation, with low values being desirable. In the present study, the prediction probability of the CatBoost model was the closest to the true probability compared with other algorithms, with a Brier score of 0.157 (see [Supplementary material online, Figure S6](#)). DCA can assess whether it is beneficial to use a model to make clinical decisions by calculating a clinical 'net benefit'. Compared with other algorithms, the CatBoost model presented the best decision benefit (see [Supplementary material online, Figure S7](#)).

Accordingly, given the above results, the CatBoost algorithm was selected as the primary ML algorithm to perform the remaining analyses.

The HPO was applied to improve the prediction performance of the CatBoost model. After 100 trials, the optimal combination of hyperparameters with the best prediction performance was obtained (see [Supplementary material online, Figure S8](#)). The final settings of the hyperparameter search are listed in [Supplementary material online, Table S3](#). The compact model, with the optimal combination of parameters, demonstrated a significantly improved discrimination ability, as indicated by the high AUROC value of 0.929 (see [Supplementary material online, Figure S9](#)).

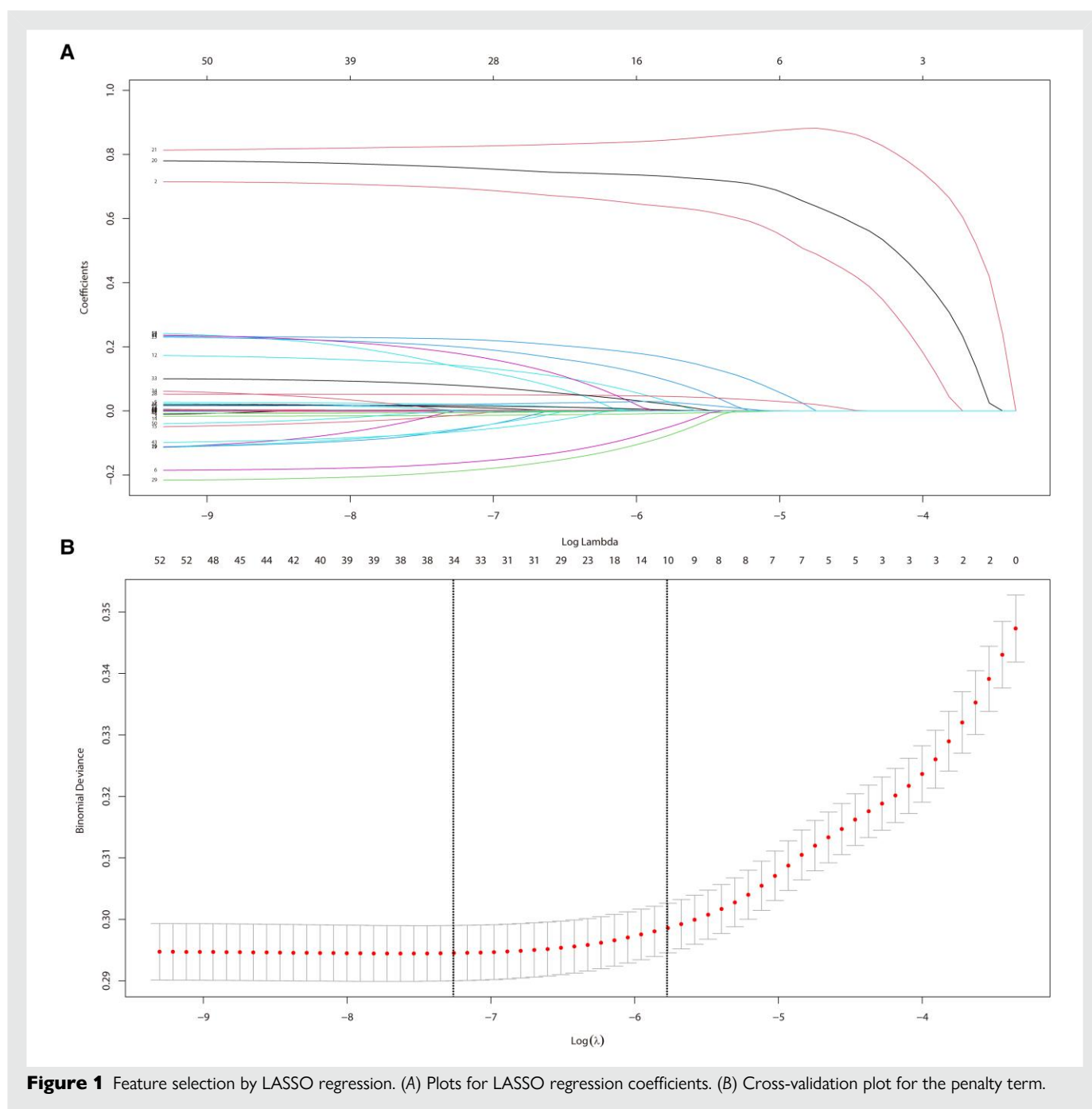
Model evaluation and validation

To further demonstrate the performance of the final model, the ROC curve was used to predict 1-year mortality based on the CatBoost model. The result showed that the CatBoost model was also able to accurately predict in-hospital mortality in sepsis patients with an AUROC of 0.837, which further demonstrated the satisfactory prediction performance of the CatBoost model (see [Supplementary material online, Figure S10](#)). External validation was conducted in the present study to demonstrate the generalization of the CatBoost model. Data of the external validation cohort were collected from the eICU database, and only the 10 key features were included in this cohort. We found that the CatBoost model presented with the greatest prediction performance compared with other ML algorithms in the validation cohort (see [Supplementary material online, Figure S11](#)), with AUROC, accuracy, sensitivity, and specificity of 0.836, 0.792, 0.699, and 0.863, respectively (see

Table 1 Baseline characteristic

Variables	Training cohort (n = 27 139)			Validation cohort (n = 9492)		
	LTVA (n = 1136)	Non-LTVA (n = 26 003)	P value	LTVA (n = 177)	Non-LTVA (n = 9315)	P value
Temperature, °C	36.7 ± 0.7	36.9 ± 0.6	<0.001	36.6 ± 1.3	36.9 ± 1.0	<0.001
SBP, mmHg	85.0 ± 17.3	89.5 ± 16.8	<0.001	95.9 ± 22.5	99.4 ± 20.9	<0.001
CCI	6.23 ± 2.83	5.65 ± 2.88	<0.001	5.98 ± 2.96	4.21 ± 2.23	<0.001
MV, %	710 (62.5%)	10 864 (41.8%)	<0.001	104 (58.8%)	2567 (27.6%)	<0.001
qSOFA score	0.78 ± 0.78	0.69 ± 0.74	<0.001	1.91 ± 0.72	1.78 ± 0.75	0.023
CHF, %	724 (63.7%)	8467 (32.6%)	<0.001	38 (21.5%)	854 (9.2%)	<0.001
AF, %	579 (51.0%)	9181 (35.3%)	<0.001	37 (20.9%)	1088 (11.7%)	<0.001
AMI, %	390 (34.3%)	2219 (8.5%)	<0.001	35 (17.8%)	302 (3.2%)	<0.001
Pneumonia, %	533 (46.9%)	9673 (37.2%)	<0.001	72 (40.7%)	3117 (33.5%)	0.044
AKI, %	765 (67.3%)	12 760 (49.1%)	<0.001	72 (40.7%)	2552 (27.4%)	<0.001

LTVA, life-threatening ventricular arrhythmia; SBP, systolic blood pressure; CCI, Charlson comorbidity index; MV, mechanical ventilation; qSOFA, quick Sequential Organ Failure Assessment; CHF, congestive heart failure; AF, atrial fibrillation; AMI, acute myocardial infarction; AKI, acute kidney injury.



Supplementary material online, Table S3). Therefore, we suggested that the CatBoost model had excellent robustness, reliability, and generalization ability, and could be promoted in clinical practice.

Additionally, we developed a personalized nomogram to visually present the prediction result, together with the probability of LTVA within hospitalization. We obtained the LTVA probabilities of each participant in the external cohort using the 'predict_proba' function of the CatBoost algorithm. Consequently, all patients in the external validation set were ranked by the prediction probability. The prediction distribution plots of the CatBoost model with patients sorted in the order of risk showed positive clustering of patients who suffered from LTVA, suggesting the favourable discriminatory ability of the model (Figure 3A). To further optimize the risk stratification, the decision tree was performed in this study. Two cut-off values of the total

points of the nomogram were calculated using the decision tree algorithm (Figure 3B). Accordingly, participants were divided into three groups: low-risk (total points < 312), middle-risk (total points ≥ 312 and < 362), and high-risk groups (total points ≥ 362) (Figure 3C). The patients in the middle-risk group had about a four-fold risk of LTVA compared with the low-risk group, and the risk was eight-fold in the high-risk group, which further demonstrated the ability of risk stratification of the nomogram (Figure 3D).

Discussion

In the present study, we built an ML-based ensemble model that enabled the accurate prediction of LTVA in patients with sepsis during

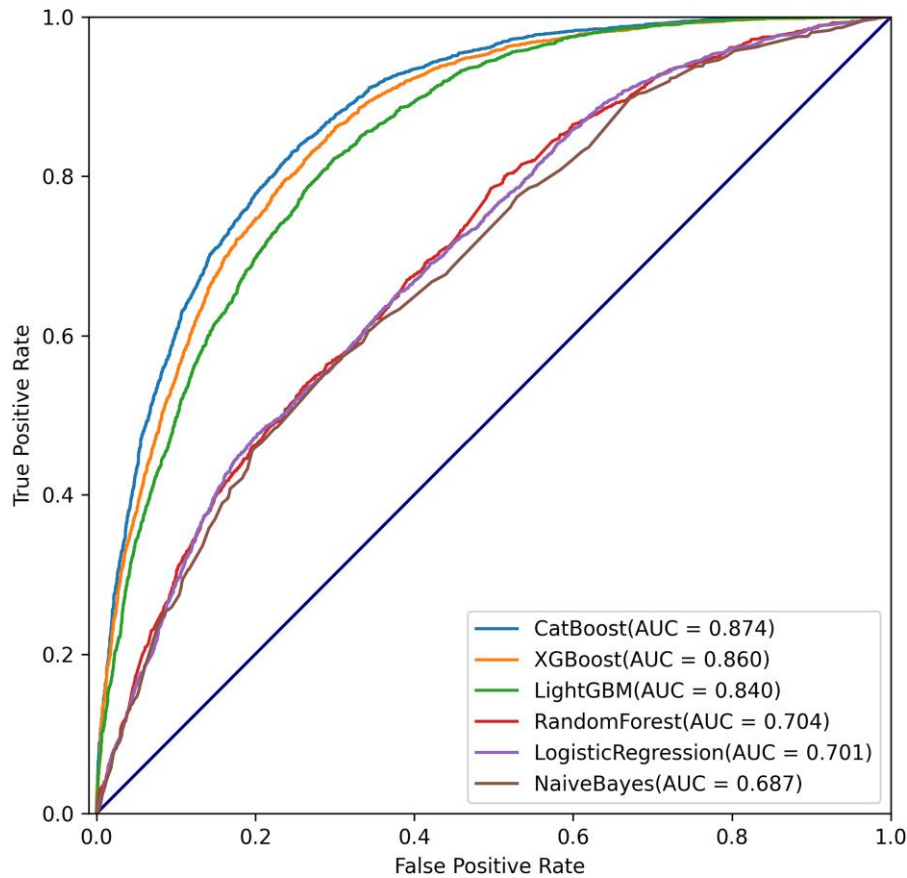


Figure 2 Prediction performance of different models.

Table 2 Internal and external validation

	AUROC	ACC	SENS	SPEC	PPV	NPV	MCC	F1-score
Internal validation								
CatBoost	0.874	0.794	0.710	0.849	0.755	0.818	0.566	0.732
XGBoost	0.860	0.791	0.680	0.846	0.743	0.802	0.536	0.710
LightGBM	0.840	0.757	0.637	0.835	0.716	0.778	0.483	0.674
Random forest	0.704	0.647	0.446	0.778	0.568	0.682	0.237	0.500
Logistic regression	0.701	0.660	0.421	0.816	0.600	0.683	0.259	0.495
Naive Bayes	0.687	0.625	0.391	0.777	0.535	0.661	0.182	0.453
External validation								
CatBoost	0.836	0.792	0.699	0.863	0.796	0.790	0.574	0.745
XGBoost	0.817	0.753	0.662	0.823	0.741	0.761	0.493	0.699
LightGBM	0.812	0.731	0.636	0.804	0.713	0.743	0.448	0.672
Random forest	0.697	0.617	0.449	0.746	0.575	0.639	0.205	0.504
Logistic regression	0.686	0.624	0.357	0.817	0.583	0.639	0.196	0.443
Naive Bayes	0.679	0.594	0.325	0.789	0.527	0.617	0.129	0.402

AUROC, area under the receiver operating characteristic curve; ACC, accuracy; SENS, sensitivity; SPEC, specificity; PPV, positive prediction value; NPV, negative prediction model; MCC, Matthews correlation coefficient.

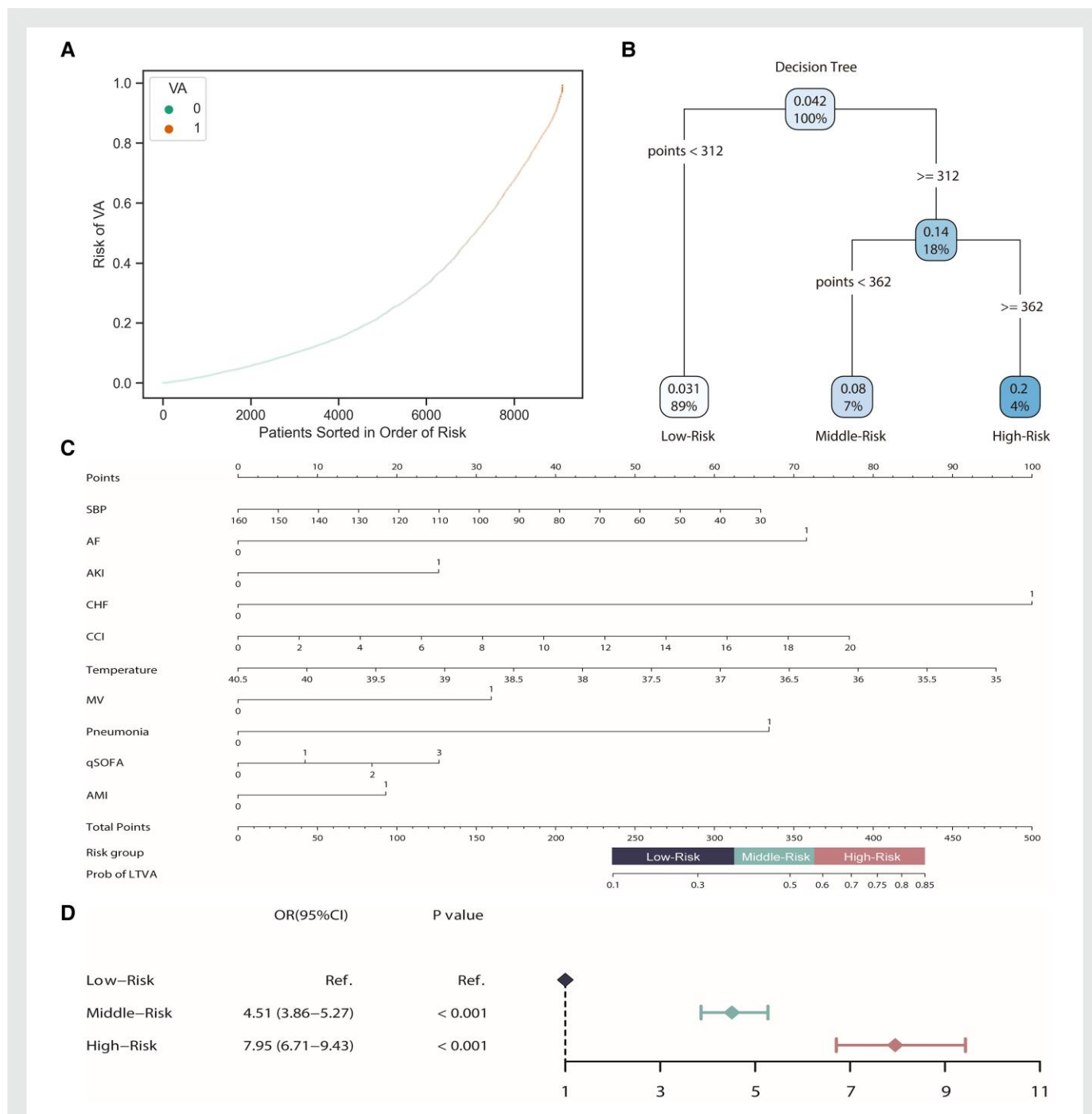


Figure 3 Model validation and risk stratification. (A) Prediction distributions of the risk of LTVA. (B) The risk stratification of LTVA according to a decision tree. (C) A nomogram of the CatBoost model for predicting LTVA during hospitalization in sepsis patients. (D) Logistic analysis of the risk of LTVA based on the risk stratification.

hospitalization. We investigated several ML algorithms to predict LTVA and found that CatBoost presented the best performance. Temperature, SBP, CCI, qSOFA score, MV, AF, AMI, CHF, pneumonia, and AKI were selected as the key variables based on feature importance to develop the prediction model. The excellent prediction performance of the model was validated both internally and externally. The prediction model may facilitate more responsive clinical care that is beneficial for sepsis patients with high-risk LTVA through early identification.

ML algorithms

ML algorithms can facilitate information processing of complex and mass data, and make reasonable interpretations through continuous learning and training, to assist clinicians to make better medical decisions. Accordingly, in recent years, numerous ML algorithms are widely employed to develop clinical prediction models including regression and classification. In this study, we built a classification prediction model

aiming to early identify sepsis patients with a high risk of LTVA during hospitalization, using six commonly used ML classification algorithms including XGBoost, LightGBM, and LR. The result showed that the CatBoost presented with the best prediction performance compared with other algorithms. CatBoost belongs to gradient boosting algorithms, can successfully handle categorical features and missing values automatically, and takes advantage of dealing with them during training instead of pre-processing time.²⁷

Life-threatening ventricular arrhythmia predictors in sepsis

A total of 10 commonly available features were employed to build the prediction model. Predictors can be considered in three categories: (i) hospitalization information of individuals, such as CCI and qSOFA score; (ii) vital signs, such as SBP and temperature; (iii) comorbidities, such as AF, CHF, AMI, pneumonia, and AKI. The 10 predictors did not involve many laboratory tests or imagological examinations, making it easy to early identify sepsis patients with a high risk of LTVA. According to its SHAP values, with the advancing of CCI and qSOFA score, the probability of LTVA is increasing as expected, because they reflect worse general health status. A recent study has found that arrhythmias causing haemodynamic compromise were more common in the hypothermia group than in the normothermia group in patients with cardiac arrest.²⁸ Electrolyte disturbances, fluid status, and a temperature effect on cardiac myocytes may help to explain the finding.²⁹ Additionally, hypothermia was found highly correlated with poor outcomes in sepsis patients.³⁰ We suggested that hypothermia was conducive to predicting LTVA in patients with sepsis which may explain the high mortality in sepsis patients with hypothermia. Pneumonia was an independent risk factor for arrhythmias and even cardiac arrest, the mechanisms may involve myocardial ischaemia, a maladaptive response to hypoxia, and sepsis-related cardiomyopathy.³¹ There is an intimate interaction between the heart and kidney, AKI can exacerbate heart dysfunction via volume and pressure overload and can aggravate cardiomyopathy due to uraemic toxins and acidaemia, and result in heart failure and LTVA.³² AF, CHF, and AMI, can significantly influence mechanical and electrical activity of the heart, and consequently, cause LTVA as expected.

Clinical application value

First, the CatBoost model had satisfactory prediction value with excellent discrimination and calibration. It is the determining factor for a clinical application whether a model could accurately predict target outcomes. Second, practicality is one of the main advantages of the model. The 10 features for prognosis prediction were readily accessible and frequently monitored in routine clinical practice; therefore, the model could be generalized on a large scale, especially for undeveloped regions. Third, the CatBoost model had certain interpretability. As we mentioned above, features contributing to LTVA risk prediction in this study were tangible and many of them had been proven intimately correlated with LTVA in sepsis patients. We introduced the SHAP method which can provide a visual interpretation of the optimal model at the global and the local levels.²² SHAP values of each feature for all patients were separately aggregated and averaged to investigate what features were considered important. Finally, other strengths of the CatBoost model include its generalizability and robustness. To verify whether the prediction performance of the model was consistent across the different periods, different regions, or different populations, and then to improve the credibility of the model, we performed external validation in the present study and found that the prediction performance of the CatBoost model was also excellent in the eICU database. Given the above characteristics, the CatBoost model, therefore, has great clinical application value in the early identification of sepsis patients who are at high risk of LTVA.

Future research direction

We have demonstrated that the CatBoost model could accurately predict LTVA during hospitalization in sepsis patients. Because patients who suffered from LTVA are at high risk of mortality, early identification of LTVA may be conducive to risk stratification, clinical decision-making, and improvement of prognosis in patients with sepsis. For instance, more intensive monitoring of vital signs, strict management of electrolytes including potassium and magnesium, and prophylactic anti-arrhythmic therapy before the sepsis is effectively contained should be applied to prevent LTVA in high-risk patients. Therefore, we plan to conduct a similar prospective study of prediction-based LTVA monitoring in sepsis patients to reduce LTVA and improve outcomes.

Limitations

Although we develop and validate a novel ML-based prediction model to accurately predict LTVA in sepsis patients, some limitations should be acknowledged. First, our work is based on a retrospective analysis of data, and further prospective studies are needed to confirm the findings. Second, data of this study were extracted from two public databases, and many important variables concerning LTVA were excluded for the high rate of missing values, such as cardiac troponin, brain natriuretic peptide, and left ventricular ejection fraction, which may put some effects on this model. Moreover, patients with CHF were not divided into reduced or preserved ejection fractions, and the prognosis between these patients was different. Third, treatments including antibiotics, vasoactive agents, or MV were not included in this study for inadequate data, which may provide bias. Fourth, variables for the models were collected at baseline. Implementation of dynamic changes of clinical markers is likely to improve prediction. In addition, the present analysis does not allow to conclude that the algorithm is specific for prediction of LTVA rather than a marker for overall poor outcome.

Conclusion

The CatBoost model described in the present study could accurately predict LTVA during hospitalization based on 10 commonly used features in patients with sepsis. The model was validated in another large database, and the results showed excellent prediction performance as well. The model could early identify patients at high risk of LTVA in sepsis and may improve outcomes by providing reasonable alerting and feedback.

Supplementary material

[Supplementary material](#) is available at *European Heart Journal – Digital Health*.

Funding

This study was supported by the Medical and Health Technology Innovation Project of Chinese Academy of Medical Sciences (2021-CXGC09-1).

Conflict of interest: The authors declare that they have no competing interests.

Data availability

The data will be shared upon reasonable request to the corresponding author.

References

1. Cecconi M, Evans L, Levy M, Rhodes A. Sepsis and septic shock. *Lancet* 2018;**392**:75–87.

2. Adhikari NKJ, Fowler RA, Bhagwanjee S, Rubenfeld GD. Critical care and the global burden of critical illness in adults. *Lancet* 2010;**376**:1339–1346.
3. Leoni D, Rello J. Cardiac arrest among patients with infections: causes, clinical practice and research implications. *Clin Microbiol Infect* 2017;**23**:730–735.
4. Huiikuri HV, Castellanos A, Myerburg RJ. Sudden death due to cardiac arrhythmias. *N Engl J Med* 2001;**345**:1473–1482.
5. Shahreyar M, Fahhoum R, Akinseye O, et al. Severe sepsis and cardiac arrhythmias. *Ann Transl Med* 2018;**6**:6.
6. Ou SM, Chu H, Chao PW, et al. Long-term mortality and major adverse cardiovascular events in sepsis survivors. A nationwide population-based study. *Am J Respir Crit Care Med* 2016;**194**:209–217.
7. Link MS, Berkow LC, Kudenchuk PJ, et al. Part 7: adult advanced cardiovascular life support: 2015 American Heart Association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. *Circulation* 2015;**132–18**(Suppl 2): S444–S464.
8. Yannopoulos D, Bartos JA, Aufderheide TP, et al. The evolving role of the cardiac catheterization laboratory in the management of patients with out-of-hospital cardiac arrest: a scientific statement from the American Heart Association. *Circulation* 2019;**139**:e530–e552.
9. Yannopoulos D, Bartos J, Raveendran G, et al. Advanced reperfusion strategies for patients with out-of-hospital cardiac arrest and refractory ventricular fibrillation (ARREST): a phase 2, single centre, open-label, randomised controlled trial. *Lancet* 2020;**396**:1807–1816.
10. Bartos JA, Grunau B, Carlson C, et al. Improved survival with extracorporeal cardiopulmonary resuscitation despite progressive metabolic derangement associated with prolonged resuscitation. *Circulation* 2020;**141**:877–886.
11. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* 2016;**375**:1216–1219.
12. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;**25**:65–69.
13. Al'aref SJ, Anchouche K, Singh G, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J* 2019;**40**:1975–1986.
14. Feeny AK, Rickard J, Patel D, et al. Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines. *Circ Arrhythm Electrophysiol* 2019;**12**:e007316.
15. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV (version 2.0). PhysioNet; 2022. <https://doi.org/10.13026/7vcr-e114>
16. 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.
17. 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.
18. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* 2016;**315**:801–810.
19. Bigger JTJ. Definition of benign versus malignant ventricular arrhythmias: targets for treatment. *Am J Cardiol* 1983;**52**:47C–54C.
20. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med* 2016;**4**:30.
21. Kok BC, Choi JS, Oh H, Choi JY. Sparse extended redundancy analysis: variable selection via the exclusive LASSO. *Multivariate Behav Res* 2021;**56**:426–446.
22. 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.
23. Mayer J, Rahman R, Ghosh S, Pal R. Sequential feature selection and inference using multi-variate random forests. *Bioinformatics* 2018;**34**:1336–1344.
24. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, AK; 2019, p. 2623–2631.
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. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;**16**:321–357.
27. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada; 2018, p. 6639–6649.
28. Dankiewicz J, Cronberg T, Lilja G, et al. Hypothermia versus normothermia after out-of-hospital cardiac arrest. *N Engl J Med* 2021;**384**:2283–2294.
29. Kelly FE, Nolan JP. The effects of mild induced hypothermia on the myocardium: a systematic review. *Anaesthesia* 2010;**65**:505–515.
30. Frank SM, Fleisher LA, Breslow MJ, et al. Perioperative maintenance of normothermia reduces the incidence of morbid cardiac events. A randomized clinical trial. *JAMA* 1997;**277**:1127–1134.
31. Carr GE, Yuen TC, Mcconville JF, et al. Early cardiac arrest in patients hospitalized with pneumonia: a report from the American Heart Association's get with the guidelines-resuscitation program. *Chest* 2012;**141**:1528–1536.
32. Rangaswami J, Bhalla V, Blair JEA, et al. Cardiorenal syndrome: classification, pathophysiology, diagnosis, and treatment strategies: a scientific statement from the American Heart Association. *Circulation* 2019;**139**:e840–e878.