A Retrospective Study of Mortality for Perioperative Cardiac Arrests Towards Personalized Treatment

Huijie Shang
First Affiliated Hospital of Zhengzhou University

Qinjun Chu
Zhengzhou Center Hospital Affiliated to Zhengzhou University

Shasha Zheng
Shanxi Province Hospital of Traditional Chinese Medicine

Muhuo Ji
Second Affiliated Hospital of Nanjing Medical University

Jin Guo
First Affiliated Hospital of Zhengzhou University

Jianjun Yang (yjyangjj@126.com)
First Affiliated Hospital of Zhengzhou University

Haotian Ye
First Affiliated Hospital of Zhengzhou University

Research Article

Keywords:

Posted Date: February 9th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1260578/v1

Abstract

Background

Perioperative cardiac arrest (POCA) is associated with a high mortality rate. This work aims to study its prognostic factors for risk mitigation by means of care management and planning.

Methods

A database of 380,919 surgeries was reviewed and 150 POCAs were curated. The main outcome is the mortality prior to hospital discharge. Patient demographics, medical history, and clinical characteristics (anesthesia and surgery) were main features. Six ML algorithms were explored: LR, SVC, RF, GBM, AdaBoost, and VotingClassifier. The last one is an ensemble of the beginning five ones. k-fold cross-validation and bootstrapping minimized the prediction bias and variance, respectively. Explainers (SHAP and LIME) were used to interpret the predictions.

Results

The ensemble provided the most accurate and robust predictions (AUC=0.90 [95% CI, 0.78-0.98]) across various age groups. The risk factors were identified by order of importance. Surprisingly, the comorbidity of hypertension was found to have a protective effect on survival, which was reported by a recent study (Alnabelsi T. et. al. 2020) for the first time to our knowledge.

Conclusions

The validated ensemble classifier in aid of the explainers improved the predictive differentiation, deepening our understanding of POCA prognostication. It offers a holistic model-based approach for personalized anesthesia and surgical treatment.

Introduction

Perioperative cardiac arrest (POCA) is a rare but extremely serious risk event with high mortality during anesthesia and surgery, which is defined commonly as the loss of circulation prompting resuscitation with chest compressions and/or defibrillation in the operating (1, 2). The reported anesthesia related POCA incidence ranges from 0.04-8 per 10,000 administered anesthetics, while it is associated with high mortality rates varying between 20-60% immediately(3–7). Accurately predicting survival and promptly making right decisions pose a huge challenge for anesthesiologists and clinicians under uncertain and dynamic environments.

The incidence and causes of cardiac arrests related to anesthesia have been studied over the last two decades (1). Nevertheless, a comprehensive understanding of POCA and controlling the risk factors are still in infancy. (8–11) studied individual variables associated with survival of cardiac arrests by meta-analysis. One main issue about this approach is that effects are often multivariate rather than univariate, making results prone to bias. Multiple disease severity scores predicting survival have been developed as a tool for risk stratification after cardiac arrest (12–21), but usually make suboptimal predictive accuracy for a specific patient population, which should be cautiously extrapolated and applied to an individual patient in hospital.

Recently, machine learning has emerged as an effective approach to integrate multiple quantitative variables to improve accuracy of incidence predictions in medicine, with the potential to dramatically improve health care delivery (22–26). Specifically, in the research of anesthesiology and cardiac arrest, it has shown in very recent years ML is a promising method for a more comprehensive understanding of the risk factors and a supporting tool for care improvement (27–31).

For this reason, this study reports all cardiac arrests that occurred in a surgical population pre-intra-post anesthesia during about an 8-years period in one of the largest Chinese tertiary hospitals, and examines causes of mortality with ML in addition to univariate method of ANOVA. After we validate the ML models, they can be used to identify the mortality risk factors and predict survival outcome of an individual patient. The data brings more information about anesthesia and surgery apart from patients' demographics, and the ML models may offer more potential to understand and manage the procedure than traditional resuscitation algorithms (33). This study will be applied to designing model-based prediction and care management strategies of anesthesia/surgery to improve the prognosis and survival of POCA.

Methods

Data collection

This retrospective study was approved by human research ethics committee of the 1st affiliated hospital of Zhengzhou university (number:KY-2021-0084). The study was registered at the Chinese clinical trial registry (ChiCTR2100051737). The requirement for obtaining written informed consent from patients was waived due to its retrospective nature. The study was performed according to the principles of the Declaration of Helsinki. Totally, 380,919 patients' electronic medical records were reviewed, who underwent a surgical procedure during the period December 2012 to June 2020. Brain-dead organ donors and babies undergoing cardiac compressions due to arrest immediately after caesarean section were excluded from the analysis. Patients on cardiopulmonary bypass, or extracorporeal membrane oxygenation, were also excluded because cardiac compressions are not needed in such situations and the use of such devices can significantly affect clinical outcomes. Amongst the anesthetic records, 150 patients who suffered POCA with a full record were enrolled in this study. Data were classified into patient demographics and operative variables. The former class included gender, age, BMI, co-morbidity

diseases, emergency, trauma, and five-category physical status by ASA PS. The latter class was comprised of anesthetic type, surgical type, operative position, the amount of blood lost and blood transfused, anti-arrhythmic drug use and continuous infusion of inotrope or vasopressor. In addition, some descriptive variables of cardiac arrest were also included, i.e., arrest time, explicit causes with operation, defibrillation situation, and duration of CPR. The primary outcome was the in-hospital mortality of POCA patients until hospital discharge.

Statistical analysis

Patient characteristics were compared by mortality outcome. The statistical methods used in this work were the same as in (31). The analyses were done in R programming languages, version 3.6.1. The code has been uploaded (refer to supplementary document 1.2 in Online Supplementary Materials).

Machine learning models

Totally, six algorithms were explored. Five of them are: LR, SVC, RF, GBM and AdaBoost, which are the most commonly used for binary classification problems in medicine (37). An additional one is an ensemble approach, which is realized through a voting classifier aggregating the prediction of multiple classifiers. Therefore, we designed a VotingClassifier, which combines the predictions of aforementioned five models in order to improve prediction robustness.

Note that it was quite a challenge to get a robust and accurate ML model, given that the data were scarce because a POCA was a very rare incidence. A thorough effort was made in this work as follows.

1) 5-fold cross-validation resampling procedure was used to evaluate the models on the limited training data to reduce the prediction bias. And a bootstrapping method was further leveraged to minimize the potentially large prediction variance. For each fold we extracted the true positive rate and false positive rate, and calculated the AUC, the mean of which was used as the optimization metric. Based on this series of results we obtained a confidence interval of AUC to show the robustness of a ML classifier.

2) Grid and random hyper-parameter searches were used to search for optimal hyperparameters.

Model explainability

The ML models except LR are all “black-box” algorithms. To break down the black box, we employed several model-agnostic methods, including 1) Permutation feature importance to globally understand the importance and effects of features ; 2) SHAP to calculate local feature importance for every observation (38); 3) LIME to analyze individual predictions (accumulated local effects) (39). All ML analyses were conducted using open-source software libraries of Python version 3.7.3.

Results

Patient characteristics and statistical analysis

As shown in Table 1, a cohort of 150 POCA patients were investigated, of which 81 patients died prior to hospital discharge, resulting in a survival of 46%. The average age was 49.4 (±18.5) years, with 96 (64.0%) being male and 73 (48.7%) being emergency. 145 (96.7%) underwent general anesthesia, and 91 (60.7%) were in ASA PS III-V. 14 patients underwent the cardiac arrest during induction, while 13 patients were in intubation. The majority of cardiac arrests (N = 102, 68.0%) occurred during the operations of surgery. The common causes of POCA were preoperative complications (N =34, 22.7%), anesthesia related (N=23, 15.3%), surgical complications (N = 41, 27.3%).

Table 1
Patient demographics and operative variables of entire cohort stratified by survival to hospital discharge.

	All patients	Survived to hospital discharge		*p*-Value
		Yes	No	
Number of patients (%)	150 (100.0)	69 (46.0)	81 (54.0)	
Gender, N (%)				0.016
Female	54 (36.0)	32 (46.4)	22 (27.2)	
Male	96 (64.0)	36 (57.2)	60 (74.1)	
Age, years (SD)	49.4 (18.5)	51.3 (19.4)	47.1 (17.2)	0.155
BMI, kg/m^2 (SD)	24.3 (3.9)	24.6 (4.1)	23.8 (3.7)	0.233
Comorbidities and medical history, N (%)				
Diabetes	10 (6.7)	7 (10.1)	3 (3.7)	0.186
Hypertension	43 (28.7)	24 (34.8)	19 (23.5)	0.146
Cardiac disease	34 (22.7)	15 (21.7)	19 (23.5)	1.000
Pulmonary disease	36 (24.0)	14 (20.3)	22 (27.2)	0.489
Hepatic disease	14 (9.3)	4 (5.8)	10 (12.3)	0.298
Renal disease	20 (13.3)	9 (13.0)	11 (13.6)	1.000
Neurological disease	31 (20.7)	13 (18.8)	18 (22.2)	0.823
Cancer	33 (22.0)	15 (21.7)	18 (22.2)	1.000
Surgical type, N (%)				0.020
Abdominal	63 (42.0)	28 (40.6)	35 (43.2)	
Neurosurgery	17 (11.3)	3 (4.3)	14 (17.3)	
Thoracic	37 (24.7)	15 (21.7)	22 (27.2)	
Throat	12 (8.0)	8 (11.6)	4 (4.9)	
Others	21 (14.0)	14 (20.3)	7 (8.6)	
Emergency, N (%)	73 (48.7)	24 (34.8)	49 (60.5)	0.005
Trauma, N (%)	19 (12.7)	11 (15.9)	8 (9.9)	0.352

NA*: not available

	All patients	Survived to hospital discharge		*p*-Value
		Yes	No	
Anaesthetic type, N (%)				1.000
General	145 (96.7)	66 (95.7)	79 (97.5)	
Local	5 (3.3)	2 (2.9)	3 (3.7)	
Operative position (%)				0.014
Left lateral decubitus	10 (6.7)	6 (8.7)	4 (4.9)	
Lithotomy	4 (2.7)	4 (5.8)	0 (0.0)	
Prone	3 (2.0)	3 (4.3)	0 (0.0)	
Right lateral decubitus	11 (7.3)	7 (10.1)	4 (4.9)	
Supine	122 (81.3)	48 (69.6)	74 (91.4)	
ASA PS, N (%)				0.000
1	4 (2.7)	4 (5.8)	0 (0.0)	
2	55 (36.7)	35 (50.7)	20 (24.7)	
3	37 (24.7)	16 (23.2)	21 (25.9)	
4	36 (24.0)	12 (17.4)	24 (21.0)	
5	18 (12.0)	1 (1.4)	17 (24.6)	
Arrest time, N (%)				0.937
Induction	14 (9.3)	7 (10.1)	7 (8.6)	
Intubation	13 (8.7)	5 (7.2)	8 (9.9)	
Surgery	102 (68.0)	46 (66.7)	56 (69.1)	
NA*	21 (14.0)	10 (14.5)	11 (13.6)	
Defibrillate, N (%)	72 (48.0)	30 (43.5)	42 (51.9)	0.482
Arrest cause, N (%)				0.200
Anesthesia	23 (15.3)	11 (15.9)	12 (14.8)	
Comorbidities	34 (22.7)	10 (14.5)	24 (29.6)	
Surgery	41 (27.3)	20 (29.0)	21 (25.9)	

NA*: not available

	All patients	Survived to hospital discharge		*p*-Value
		Yes	No	
Unknown	52 (34.7)	27 (39.1)	25 (30.9)	
Hemorrhage, median [Q1, Q3] (ml)	100.0 [2.3, 500.0]	95.0 [4.3, 200.0]	200.0 [0.8, 1075.0]	0.032
Blood transfusion, median [Q1, Q3] (ml)	0.0 [0.0, 1000.0]	0.0 [0.0, 0.0]	0.0 [0.0, 1712.0]	0.002
Epinephrine, median [Q1, Q3] (mg)	2.0 [0.1, 5.9]	0.5 [0.0, 2.0]	4.0 [2.0, 7.9]	0.000
Atropine, median [Q1, Q3] (mg)	0.0 [0.0, 0.5]	0.0 [0.0, 0.5]	0.0 [0.0, 0.5]	0.737
Amiodarone, median [Q1, Q3] (g)	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.598
Ephedrine, median [Q1, Q3] (mg)	0.0 [0.0, 0.0]	0.0 [0.0, 2.3]	0.0 [0.0, 0.0]	0.182
Methoxamine, median [Q1, Q3] (mg)	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0,0.0]	0.885
CPR, median [Q1, Q3] (min)	30.0 [10.0, 37.0]	11.0 [1.0, 37.0]	37.0 [27.0, 43.0]	0.000
NA*: not available				

On the one hand, the following variables were statistically different between the survivor and non-survivor groups ($p<0.05$): gender, surgical type, emergency, operative position, ASA PS, hemorrhage, blood transfusion, epinephrine, and CPR. Accordingly, it could be induced that the favorable variables for survival were being female, throat (or others) surgery, none-emergency, and ASA PS I-II. On the contrary, a higher probability of mortality occurred in male, neurosurgery, emergency, operative position of supine, massive hemorrhage and blood transfusion, and ASA PS V. A higher epinephrine dose (4.0 [IQR 2.0-7.9] versus 0.5 [IQR 0.0-2.0] mg) was administrated and a longer CPR (37.0 [IQR 27.0-43.0] versus 11.0 [IQR 1.0-37.0] min) were performed during cardiac arrest in non-survivors.

On the other hand, other variables were not found significantly different between the two groups, such as age, BMI, trauma, arrest time, defibrillate, and arrest cause. And there was no evidence that the administration of other drugs except epinephrine was directly associated with survival or death.

In addition, the comorbidities and medical history generally did not show to be strongly linked with the morality, while the observed difference regarding hypertension ($p=0.146$) between the survival and death groups was attributed to chance only by 14.6%. It indicates that hypertension might be one remarkably influential comorbidity for further exploration.

ML models

The 150 patients were split into two subgroups in a gender-stratified way, i.e., 112 (75%) and 38 (25%) for training and testing of the ML models, respectively. In order to preserve the same gender proportions of patients in each subgroup as observed in the total patients, the data was split in a gender-stratified way. The predicting outcome was the probability of mortality.

Fig. 1 is the ROC curves generated with the test data by the six machine learning models: LR, SVC, RF, GBM, AdaBoost, and VotingClassifier. Areas under the curves were 0.84, 0.87, 0.91, 0.90, 0.87 and 0.90 for the six models, respectively.

In binary classification, the most basic metric/bench-mark is the confusion matrix as “accuracy”, “precision”, “recall”, “f1-score”, “ROC” and “AUC” all stem from the confusion matrix (38). We used these multi-perspective performance measures to fairly judge the predictive models.

Table 2

Performance of the six ML models for the estimation of mortality of patients with a POCA. The 95% CI of AUC was calculated from 1000 bootstrap resamples of predictions on the test data.

Models	AUC [95%CI]	accuracy	precision	recall	f1-score
Logistic regression	0.84 [0.71-0.95]	0.74	0.78	0.70	0.74
Support vector classifier	0.87 [0.73-0.96]	0.79	0.83	0.75	0.79
Random forest	0.91 [0.79-0.98]	0.82	0.84	0.80	0.82
Gradient boost machine	0.90 [0.79-0.98]	0.82	0.84	0.80	0.82
Adaptive boosting classifier	0.87 [0.73-0.97]	0.76	0.79	0.75	0.77
Ensemble (VotingClassifer)	0.90 [0.78-0.98]	0.84	0.85	0.85	0.85

As shown by Table 2, three significantly accurate ML models were the RF (AUC, 0.91 [95% CI, 0.79-0.98]), the ensemble (AUC, 0.90 [95% CI, 0.78-0.98]), and the GBM (AUC, 0.90 [95% CI, 0.79-0.98]). It is not a surprise that as a simple and interpretable classifier, the LR produced the poorest accuracy (AUC, 0.84 [95% CI, 0.71-0.95]). Taking other metrics into account, it is demonstrated that the VotingClassifier outperformed all the other classifiers with the highest values of accuracy (0.84), precision (0.85), recall (0.85), and f1-score (0.85).

We dug further into two aspects to analyze the prediction performance of the models. One was probability curves for each ML models (Fig. 2); another was model comparisons with respect to mortality estimation across age groups (Fig. 3).

In Fig. 2, the LR estimated a higher probability of survival. Corresponding to a threshold of 50%, the false negative (FN) of mortality was six and the false positive (FP) was four, i.e., six died patients were wrongly classified into survivors while five survived patients were predicted to be died. For the SVC model, FN=5

and FP=3, with low variance in probability attributed to all survivors. For the RF and the GBM, the misclassified values were smaller, i.e., FN=4 and FP=3. And the VotingClassifer brought about the smallest misclassifications, with FN=3 and FP=3. In addition, the GBM and VotingClassifier demonstrated significant separation of the died from the survived patients, with lower overlap in the two groups.

It was further demonstrated in Supplementary Figure1 the VotingClassifier was the best classifier for age groups “<12” and “>=65”, and probably the second best for age groups “12~40” and “40~65” (outmatched only by the RF). The GBM model tended to significantly overestimate mortality in age groups “<12” and “>=65.

To summarize, the ensemble ML model (VotingClassifier) outperformed all the other classifiers, which made better predictions and achieved better performance than any single contributing model. Moreover, it reduced the spread or dispersion of the predictions with higher robustness.

Model explainability

Firstly, we applied the SHAP to explain predictions on the test data by the VotingClassifier. The SHAP summary, combing feature importance with feature effects, was visualized with violin plots to see the distribution of Shapley values (Fig. 3). The position on the y-axis was determined by the feature and on the x-axis by the Shapley value.

The following results were obtained and most of them enhanced the previous ANOVA analyses:

1) A high mortality risk was strongly associated with the top 10 important features, by order of importance, a longer CPR (>=60 min), a higher ASA PS (IV-V), surgical type (= “abdominal” or “neurosurgery”), a higher dose of epinephrine (>6mg), emergency, being male, massive hemorrhage (>=800ml), being elder (especially >65 years), cause of arrest (= “anesthesia” or “comorbidities”), or massive blood transfusion (>=800ml);

2) In the less important features, operative position (= “supine”), arrest time (= “induction”), comorbidity (= “cancer” or “hepatic disease”), BMI (= “obese”), and atropine (>0.65 mg) showed slightly positive association with the mortality;

3) Counter-intuitively but interestingly, the comorbidity of hypertension appeared to have a protective effect on survival prior to hospital discharge, a finding which was previously reported recently (33).

All effects described the model behavior and were not necessarily causal in the real world, which was why we used “association” rather than “causation” in above statement (32).

Secondly, we interpreted the VotingClassifier with the LIME explainer, particularly to explore misclassification of prediction. Four typical cases corresponding to the four quadrants in a confusion matrix: TP, TN, FP, FN, were compared. The top 10 features were presented in Fig. 4, with the weight of each feature represented in either green or red up to whether it favored survival or death.

In one TP case (Fig. 4a), we showed a specific individual with a high probability of mortality (80%). This patient died as predicted, and the key risk-associated factors were a longer CPR (30-60min, ~20% impact on mortality), ASA PS=IV-V (~14% impact), epinephrine>5mg (~12% impact), emergency (~9% impact), obese (~8% impact), and no hypertension disease (~7% impact). In one TN case (Fig. 4b), the predicted probability of mortality was 24%. The patient actually survived and was correctly predicted. The survivor-favorable features were CPR<=30min (~27% increase probability of survivor), ASA PS=I-III (~14% increase), BMI=underweight, hemorrhage <200ml, being female, and no hepatic disease.

In one FP case (Fig. 4c), the predicted probability of mortality was 62% but the patient survived. It showed that the key unfavorable features for survivor were CPR=30~60min (~19% impact) and hemorrhage >=800ml (~12% impact), to which the misclassification could be attributed. In one FN case (Fig. 4d), the predicted probability of mortality was 39%. However, the patient died. The most survivor-favorable feature was found to be CPR≤30min (~ -28% impact), which probably overshadowed other survivor-unfavorable features, such as Emergency, leading to this misclassification.

Discussion

In this study, the ML models except LR are “black-box” algorithms, providing great accuracy at the cost of low interpretability (33). The danger of decision made by ML without breaking open the black box is multiple-fold: 1) It is usually hard to explain to clinicians the predictions, which is a barrier to the adoption of ML for high stakes decisions(35); 2) More and more concerns or regulations specific to ML have been emerging on interpretability and its predictive reasoning (for example, the EU General Data Protection Regulation).

Firstly, a global model-agnostic method of permutation feature importance was employed in this work. The results were not shown in this article because some evident drawbacks of this method were found: 1) shuffling the feature added randomness and the results usually varied greatly; 2) some features were inherently correlated and this method was very biased by unrealistic data instances.

In this study, SHAP and LIME were demonstrated to be two competent local model-agnostic methods in the model explainability. Instead of calibrating global feature contributions, these two methods train local surrogate models to explained individual predictions with more solid insights generated, such as how to rank a feature by importance with a favorable or unfavorable impact value on prediction outcome. We got contrastive explanations with the two explainers, particularly to explore misclassification, making the ensemble ML model more transparent and shedding light on their applications in clinical decision-making. Explanations can be used to interrogate and rectify the ensemble model when such misclassification surfaces.

Our study has several limitations. First of all, in spite of the fact the ultimately validated ensemble model was robust and accurate, the size of data used was still relatively. Especially, there were only eight patients younger than 12 years, which was probably why most of the ML models (except the ensemble) failed to satisfactorily predict the outcome of this age group. In future, more internal data and even

external data will bring more benefits to establish generalizability and further increase model fidelity. Secondly, our dataset had no information on post-arrest care and discharge disposition. So, it was impossible to systematically follow up and assess long-term recovery and survival of discharged patients.

Last but not least, a ML model is not a “magic button”, although it would have reached a “super-human” performance. Like most ML approaches, the ML models validated in this study focused on predicting outcomes rather than understanding causality, i.e., they were finding correlations but not causation. As an example, it was revealed in this study that two top predictors of risk for in-hospital mortality were CPR and epinephrine. The ensemble model was predicting that the longer CPR and the bigger dose of epinephrine, the higher probability of death occurred. In fact, the opposite was true: patients (of severe ASA PS or massive hemorrhage) would be at a higher risk for serious complications and sequelae, even mortality, if insufficient CPR and/or epinephrine treatment were not timely delivered.

In clinical practice, accurate prediction models allow for improved medical prognostication, earlier identification of patients at high risk of complications, better risk adjustment and utilization of critical care resources, and more effective patient-physician-family communication.

In this study, the validated ensemble model provides superior prediction accuracy by virtue of high fidelity to data across various age groups and high robustness to uncertainty, as well as good discrimination between survivors and non-survivors. The data comprised operative parameters in addition to patients’ demographics, which makes it possible to integrate operational optimization and/or tactical planning with the model by managing the operative parameters and procedure.

One application scenario is early recognition of problems and suggestion of actions to avoid critical events. For an individual patient, an optimal combination of anesthesia management, surgical type, operative position (if optional), and treatment drugs could lead to a significantly improved probability of survival until hospital discharge (some exploratory simulations were done but not shown in this article). Another application scenario is that the model could enhance rational patient risk monitoring during operations, with drug doses administered in a timely fashion (target-controlled infusion) resulting in a precision, efficacy and safety of intravenous anesthesia delivery. In short, this model-based optimization opens up an avenue for a personalized anesthesia and surgery strategy, with a better treatment and a higher survival rate attained.

Furthermore, clinicians hesitate to apply a black-box algorithm that is hard for them to understand and trust (33, 34). In this work, the explainers (LIME and SHAP) will pinpoint logics of decision-making and mitigate issue of clinical liability, encouraging them to understand and leverage ML to assist decision making and change management in practice.

Conclusion

We demonstrated that the ensemble ML model makes solid predictions of mortality on the data of POCA patient demographics and operative parameters, bringing a more comprehensive understanding of the risk factors and patient prognostics prior to hospital discharge, compared to the ANOVA. Furthermore, the explainers of LIME and SHAP provide a more comprehensible and holistic approach to the assessment of prognosis of an individual patient. All these results may assist risk management of in-hospital cardiac arrest with improved patient-centered and personalized care.

Glossary

POCA= Perioperative cardiac arrest; ML= Machine Learning; SHAP = SHapley Additive exPlanations; LIME = Local Interpretable Model-agnostic Explanations; ANOVA = Analysis of Variance; BMI = Body Mass Index; CPR = cardiopulmonary resuscitation; ASA PS = American Society of Anesthesiologist's Physical Status classification; LR =Logistic Regression; SVC = Support Vector Classifier; RF = Random Forest, GBM = Gradient Boosted Machine; AdaBoost = Adaptive Boosting Classifier; AUC = area under the receiver operating characteristic curve; ROC = Receiver Operating Characteristic

Declarations

Data Availability Statement

The datasets generated and analyzed during the current study are available in the following repository *https://github.com/niuneo/Risk-factor-analysis-of-mortality-for-perioperative-cardiac-arrest-using-machine-learning.*

References

1. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-Hospital Cardiac Arrest: A Review. JAMA: the journal of the American Medical Association 2019;321:1200–10.
2. Kazaure HS, Roman SA, Rosenthal RA, Sosa JA. Cardiac arrest among surgical patients: an analysis of incidence, patient characteristics, and outcomes in ACS-NSQIP. JAMA SURG 2013;148:14–21.
3. Ellis SJ, Newland MC, Simonson JA, Peters KR, Romberger DJ, Mercer DW, Tinker JH, Harter RL, Kindscher JD, Qiu F, Lisco SJ. Anesthesia-related cardiac arrest. ANESTHESIOLOGY 2014;120:829–38.
4. Huo T, Sun L, Min S, Li W, Heng X, Tang L, Zhu S, Dong H, Wang Q, Xiong L. Major complications of regional anesthesia in 11 teaching hospitals of China: a prospective survey of 106,569 cases. J CLIN ANESTH 2016;31:154–61.
5. Jansen G, Irmscher L, May TW, Borgstedt R, Popp J, Scholz SS, Rehberg SW. Incidence, characteristics and risk factors for perioperative cardiac arrest and 30-day-mortality in preterm infants requiring non-cardiac surgery. J CLIN ANESTH 2021;73:110366.

6. Jansen G, Borgstedt R, Irmscher L, Popp J, Schmidt B, Lang E, Rehberg SW. Incidence, Mortality, and Characteristics of 18 Pediatric Perioperative Cardiac Arrests: An Observational Trial From 22,650 Pediatric Anesthesias in a German Tertiary Care Hospital. Anesthesia & Analgesia 2021;133:747–54.
7. Nunnally ME, O Connor MF, Kordylewski H, Westlake B, Dutton RP. The Incidence and Risk Factors for Perioperative Cardiac Arrest Observed in the National Anesthesia Clinical Outcomes Registry. Anesthesia & Analgesia 2015;120:364–70.
8. Sobreira-Fernandes D, Teixeira L, Lemos TS, Costa L, Pereira M, Costa AC, Couto PS. Perioperative cardiac arrests–A subanalysis of the anesthesia -related cardiac arrests and associated mortality. J CLIN ANESTH 2018;50:78–90.
9. Sprung J, Warner ME, Contreras MG, Schroeder DR, Beighley CM, Wilson GA, Warner DO. Predictors of survival following cardiac arrest in patients undergoing noncardiac surgery: a study of 518,294 patients at a tertiary referral center. ANESTHESIOLOGY 2003;99:259–69.
10. Hur M, Lee HC, Lee KH, Kim JT, Jung CW, Park HP. The incidence and characteristics of 3-month mortality after intraoperative cardiac arrest in adults. Acta Anesthesiol Scand 2017;61:1095–104.
11. Siriphuwanun V, Punjasawadwong Y, Saengyo S, Rerkasem K. Incidences and factors associated with perioperative cardiac arrest in trauma patients receiving anesthesia. Risk Manag Healthc Policy 2018;11:177–87.
12. V Subramanian, EJ Mascha, MW Kattan. Developing a Clinical Prediction Score: Comparing Prediction Accuracy of Integer Scores to Statistical Regression Models. Anesthesia & Analgesia. 2021. 132 (6), 1603–1613
13. Cooper S, Evans C. Resuscitation Predictor Scoring Scale for inhospital cardiac arrests. EMERG MED J 2003;20:6–9.
14. Balan P, Hsi B, Thangam M, Zhao Y, Monlezun D, Arain S, Charitakis K, Dhoble A, Johnson N, Anderson HV, Persse D, Warner M, Ostermayer D, Prater S, Wang H, Doshi P. The cardiac arrest survival score: A predictive algorithm for in-hospital mortality after out-of-hospital cardiac arrest. RESUSCITATION 2019;144:46–53.
15. Choi JY, Jang JH, Lim YS, Jang JY, Lee G, Yang HJ, Cho JS, Hyun SY. Performance on the APACHE II, SAPS II, SOFA and the OHCA score of post-cardiac arrest patients treated with therapeutic hypothermia. PLOS ONE 2018;13:e196197.
16. Constant A, Montlahuc C, Grimaldi D, Pichon N, Mongardon N, Bordenave L, Soummer A, Sauneuf B, Ricome S, Misset B, Schnell D, Dubuisson E, Brunet J, Lasocki S, Cronier P, Bouhemad B, Loriferne J, Begot E, Vandenbunder B, Dhonneur G, Bedos J, Jullien P, Resche-Rigon M, Legriel S. Predictors of Functional Outcome after Intraoperative Cardiac Arrest. ANESTHESIOLOGY 2014;121:482-91.
17. Ebell MH, Jang W, Shen Y, Geocadin RG. Development and Validation of the Good Outcome Following Attempted Resuscitation (GO-FAR) Score to Predict Neurologically Intact Survival After In-Hospital Cardiopulmonary Resuscitation. JAMA INTERN MED 2013;173:1872.
18. Fugate JE, Rabinstein AA, Claassen DO, White RD, Wijdicks EFM. The FOUR Score Predicts Outcome in Patients after Cardiac Arrest. NEUROCRIT CARE 2010;13:205–10.

19. Seewald S, Wnent J, Lefering R, Fischer M, Bohn A, Jantzen T, Brenner S, Masterson S, Bein B, Scholz J, Gräsner JT. CaRdiac Arrest Survival Score (CRASS) — A tool to predict good neurological outcome after out-of-hospital cardiac arrest. RESUSCITATION 2020;146:66–73.
20. Vane MF, Carmona MJC, Pereira SM, Kern KB, Timerman S, Perez G, Vane LA, Otsuki DA, Auler Jr JOC. Predictors and their prognostic value for no ROSC and mortality after a non-cardiac surgery intraoperative cardiac arrest: a retrospective cohort study. SCI REP-UK 2019;9.
21. Al'Aref SJ, Maliakal G, Singh G, van Rosendael AR, Ma X, Xu Z, Alawamlh O, Lee B, Pandey M, Achenbach S, Al-Mallah MH, Andreini D, Bax JJ, Berman DS, Budoff MJ, Cademartiri F, Callister TQ, Chang HJ, Chinnaiyan K, Chow B, Cury RC, DeLago A, Feuchtner G, Hadamitzky M, Hausleiter J, Kaufmann PA, Kim YJ, Leipsic JA, Maffei E, Marques H, Goncalves PA, Pontone G, Raff GL, Rubinshtein R, Villines TC, Gransar H, Lu Y, Jones EC, Pena JM, Lin FY, Min JK, Shaw LJ. Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the CONFIRM registry. EUR HEART J 2020;41:359–67.
22. Xue B, Li D, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. JAMA Network Open. 2021 Mar 1;4(3):e212240.doi: 10.1001/jamanetworkopen.2021.2240
23. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. N Engl J Med 2018;378:981–3.
24. Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial Intelligence in Anesthesiology. ANESTHESIOLOGY 2020;132:379–94.
25. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, Xu S, Stub D, Smith K, Tacey M, Liew D, Pilcher D, Kaye DM. Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. PLOS MED 2018;15:e1002709.
26. Harford S, Darabi H, Del Rios M, Majumdar S, Karim F, Vanden Hoek T, Erwin K, Watson DP. A machine learning based model for Out of Hospital cardiac arrest outcome classification and sensitivity analysis. RESUSCITATION 2019;138:134–40.
27. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. CRIT CARE MED 2016;44:368–74.
28. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest. J AM HEART ASSOC 2018;7.
29. Wu TT, Lin XQ, Mu Y, Li H, Guo YS. Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes. CLIN CARDIOL 2021;44:349–56.
30. Moitra VK, Einav S, Thies K, Nunnally ME, Gabrielli A, Maccioli GA, Weinberg G, Banerjee A, Ruetzler K, Dobson G, McEvoy MD, O Connor MF. Cardiac Arrest in the Operating Room. Anesthesia & Analgesia 2018;126:876–88.

31. Alnabelsi T, Annabathula R, Shelton J, Paranzino M, Faulkner SP, Cook M, Dugan AJ, Nerusu S, Smyth SS, Gupta VA. Predicting in-hospital mortality after an in-hospital cardiac arrest: A multivariate analysis. Resusc Plus 2020;4:100039.
32. P Schober, EJ Mascha, TR Vetter. Statistics From A (Agreement) to Z (z Score): A Guide to Interpreting Common Measures of Association, Agreement, Diagnostic Accuracy, Effect Size, Heterogeneity, and Reliability in Medical Research. Anesthesia & Analgesia 133 (6), 1633–1641.
33. Poon AIF, Sung JJY. Opening the black box of AI-Medicine. J GASTROEN HEPATOL 2021;36:581–4.
34. J Feldman, K Kuck, TM Hemmerling. Black Box, Gray Box, Clear Box? How Well Must We Understand Monitoring Devices? Anesthesia & Analgesia: 2021, 132(6): 1777–1780.
35. The LRM. Opening the black box of machine learning. The lancet respiratory medicine 2018;6:801.
36. Hemmerling TM. Automated anesthesia. Current Opinion in Anesthesiology. 2009, 22 (6), 757–763.
37. Hastie TJ, Tibshirani RJ, Friedman JH. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Published in 2009 in New York NY) by Springer.
38. Lundberg S and Lee SI. A unified approach to interpreting model predictions. 25 Nov 2017, arXiv:1705.07874.
39. Ribeiro MT, Singh S, and Guestrin C. Why should I trust you?: Explaining the predictions of any classifier. 9 Aug 2016, arXiv:1602.04938.

Figures

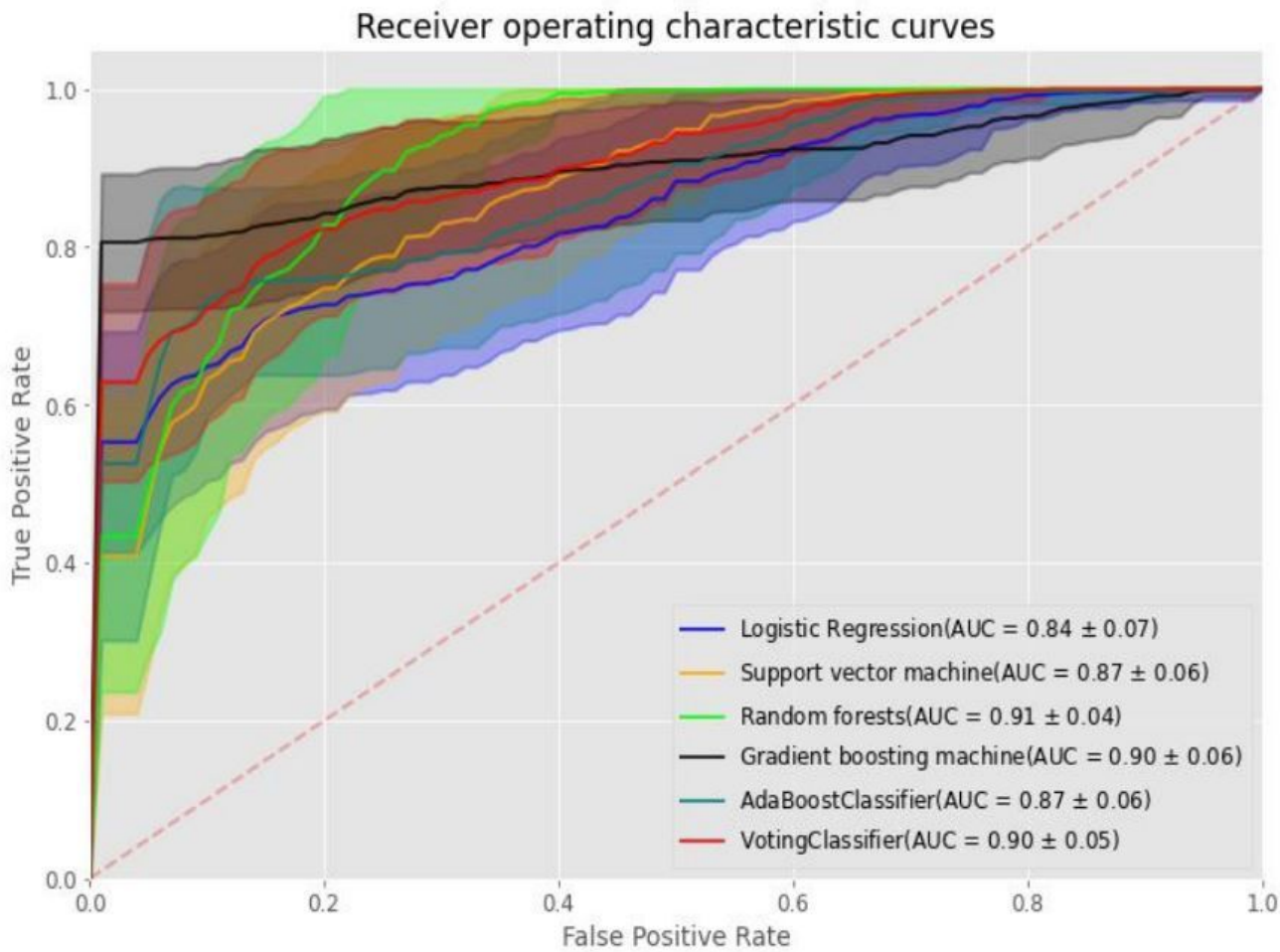

Figure 1

Figure 1

ROC curves for the six ML models on the test data. The AUC value of each model is represented by "(AUC = mean ± standard deviation)", which was estimated from 1000 bootstrap resamples of predictions on the test data. Each ROC curve is visualized by corresponding plot with shaded bands.

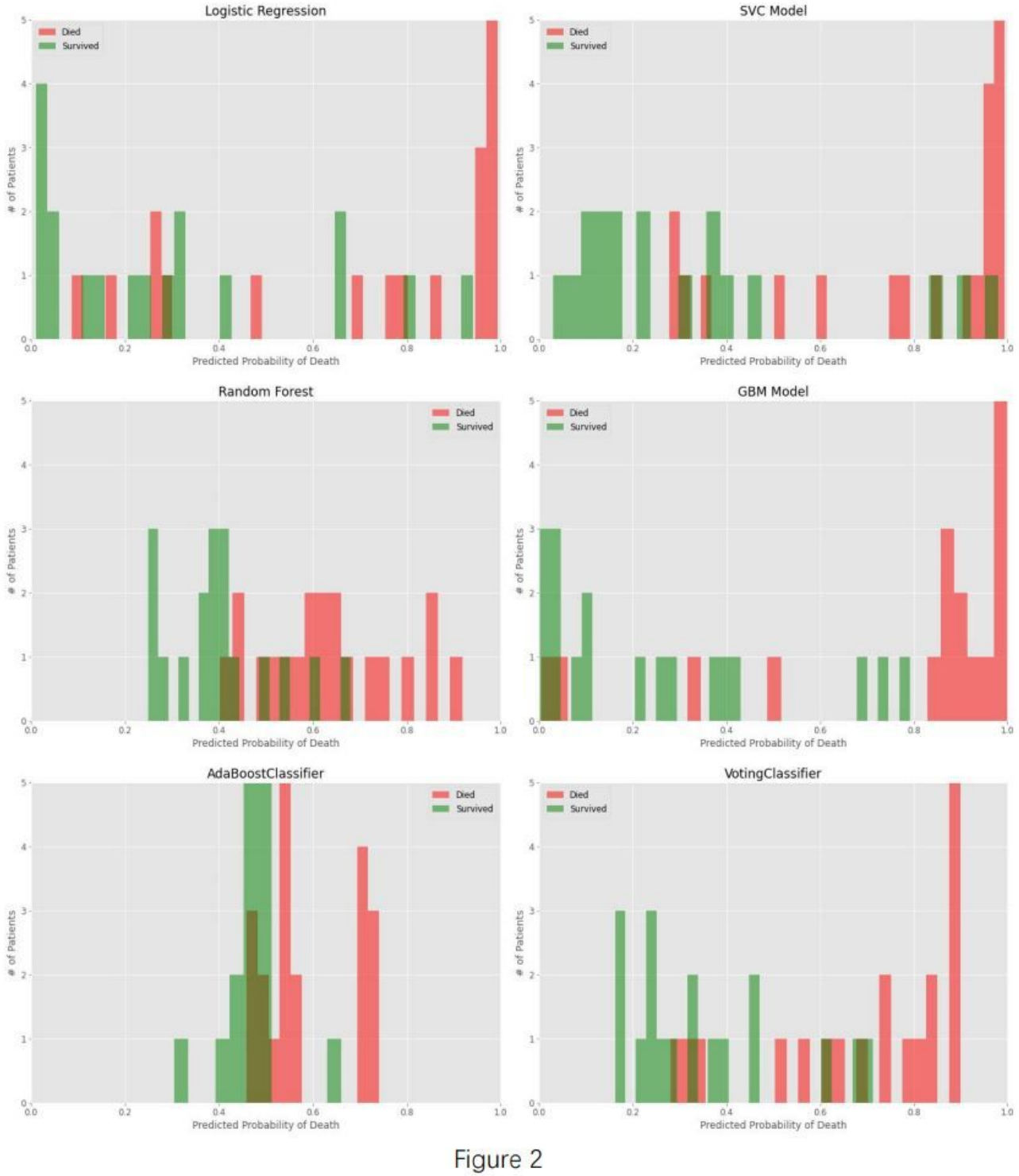

Figure 2

Probability curves for each ML model. Survivors indicated in green, and non-survivors in red. $p < 0.005$ for ensemble versus other models.

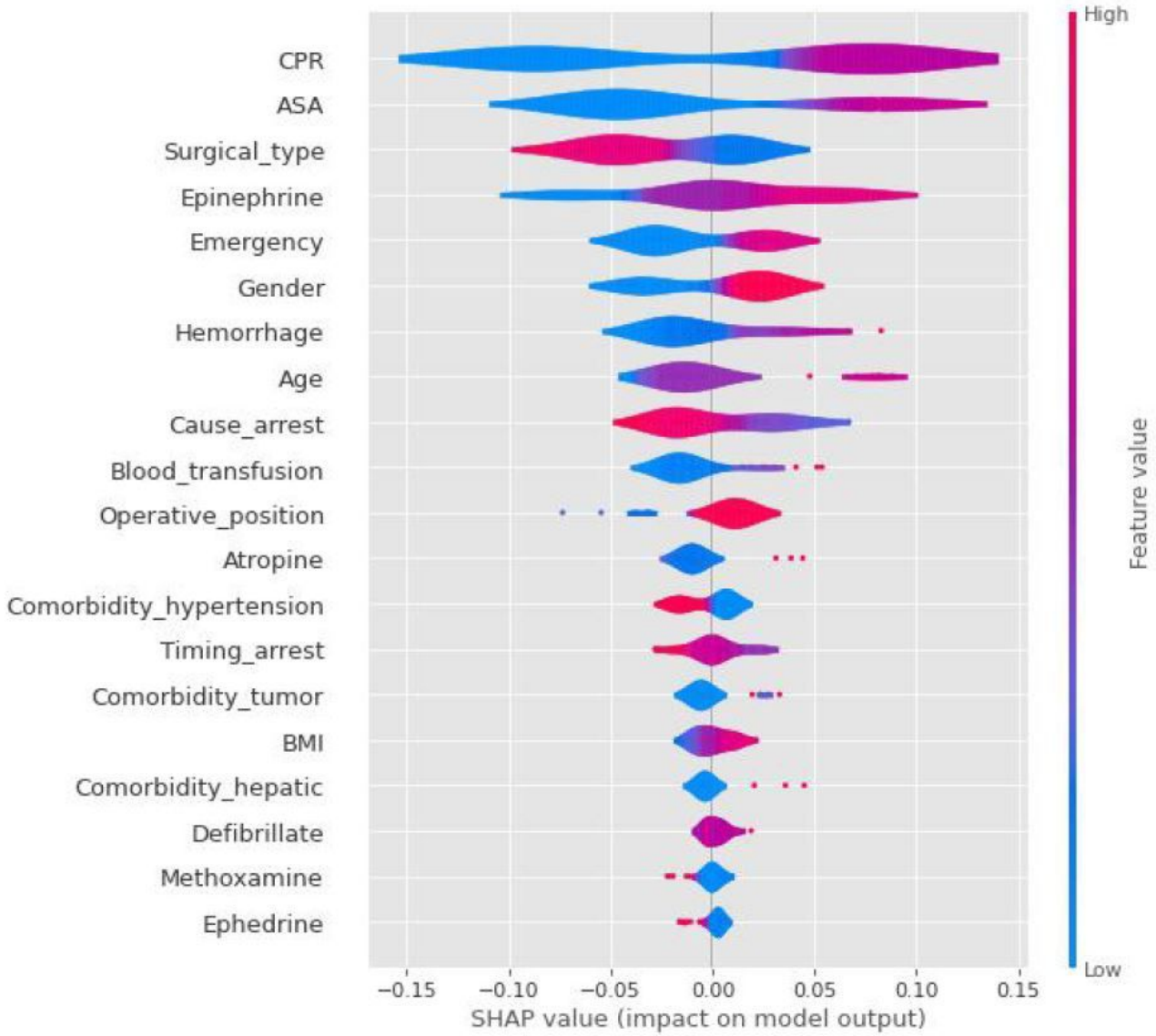

Figure 3

Figure 3

SHAP importance plots of the mortality and risk factors for the ensemble ML model (VotingClassifier). The features are ranked by importance. Each row represents the impact of a feature on the outcome of mortality, with higher SHAP values indicating higher likelihood of a positive outcome. For a binary feature, like gender, “male”→ “1” is shown in red while “female” → “0” is shown in blue. For the detailed mapping of categorical features, please refer to the code online (such as “<12 ys”→“0”, “12~40 ys”→“1”, “40~65 ys”→“2”, “>65 ys”→“3”).

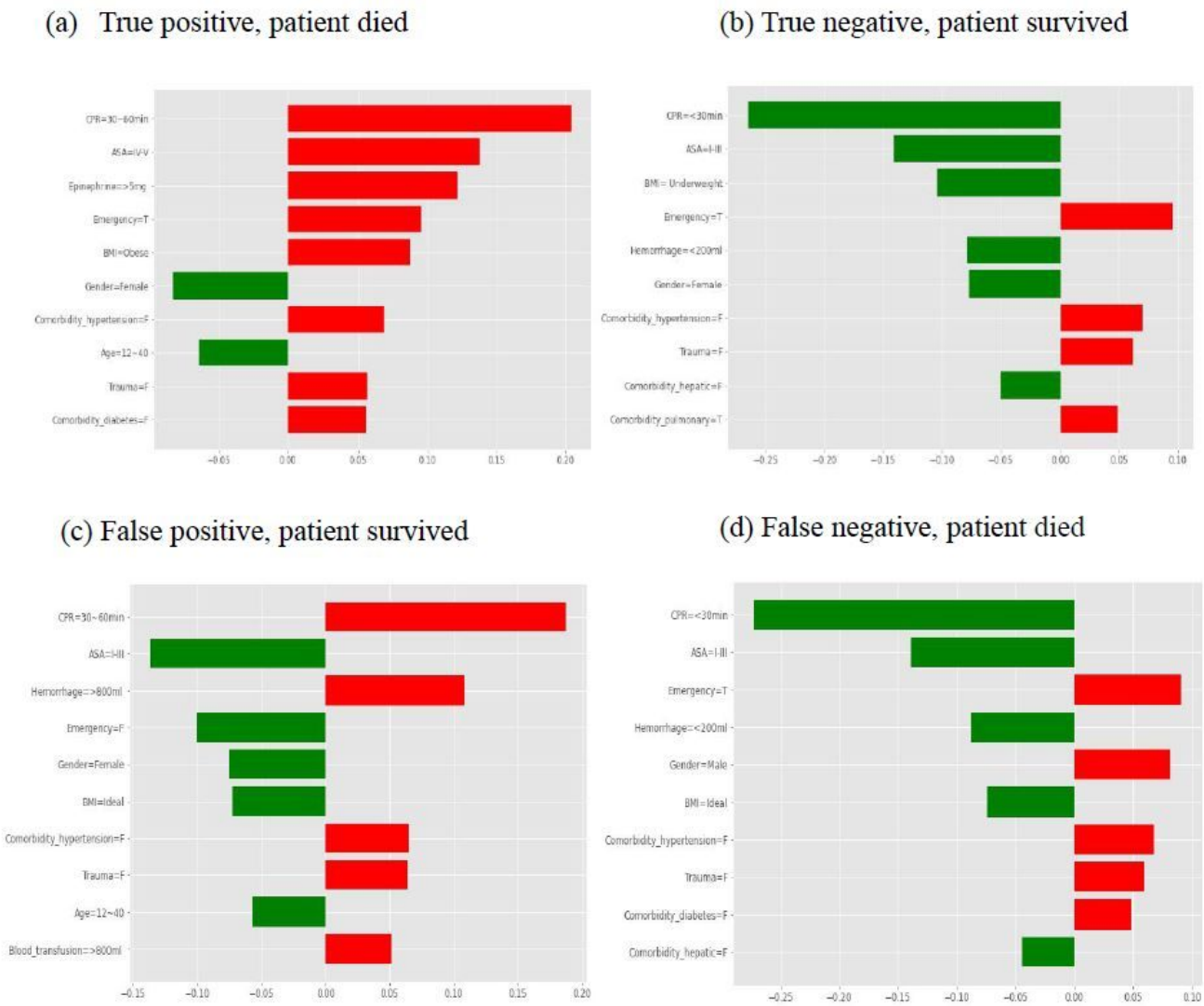

Figure 4

Figure 4

LIME explainer for four typical scenarios. (a) True positive, patient died, i.e., a correctly classified non-survivor, (b) True negative, patient survived, i.e., a correctly classified survivor, (c) False positive, patient survived, i.e., an incorrectly classified survivor (predicted to die), and (d) False negative, patient died, i.e., an incorrectly classified non-survivor (predicted to survive). Features with a green bar favored survival, and those with a red bar were predictive of mortality. The x-axis shows how much each feature added to or subtracted from the final probability value for the patient. Each weight can be interpreted in the context of the original probability; if a feature is absent for a patient, it can be numerically added to or subtracted directly from the initial probability.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineSupplementalMaterials.docx