A comparison of random survival forest and Cox regression for prediction of mortality in patients with hemorrhagic stroke

Yuxin Wang[1], Yuhan Deng[1], Yinliang Tan[1], Meihong Zhou[1], Yong Jiang[2,3*] and Baohua Liu[1*]

Abstract

Objective To evaluate RSF and Cox models for mortality prediction of hemorrhagic stroke (HS) patients in intensive care unit (ICU).

Methods In the training set, the optimal models were selected using five-fold cross-validation and grid search method. In the test set, the bootstrap method was used to validate. The area under the curve(AUC) was used for discrimination, Brier Score (BS) was used for calibration, positive predictive value(PPV), negative predictive value(NPV), and F1 score were combined to compare.

Results A total of 2,990 HS patients were included. For predicting the 7-day mortality, the mean AUCs for RSF and Cox regression were 0.875 and 0.761, while the mean BS were 0.083 and 0.108. For predicting the 28-day mortality, the mean AUCs for RSF and Cox regression were 0.794 and 0.649, while the mean BS were 0.129 and 0.174. The mean AUCs of RSF and Cox versus conventional scores for predicting patients' 7-day mortality were 0.875 (RSF), 0.761 (COX), 0.736 (SAPS II), 0.723 (OASIS), 0.632 (SIRS), and 0.596 (SOFA), respectively.

Conclusions RSF provided a better clinical reference than Cox. Creatine, temperature, anion gap and sodium were important variables in both models.

Keywords Hemorrhagic stroke, Random survival forest, Cox regression, Intensive care unit

Introduction

Stroke is the leading cause of death and long-term disability worldwide [1]. 2019 global burden of disease study (GBD) data [2]shows that stroke remains to be the second leading cause of death (11.6% of deaths) and the third leading cause of disability (5.7% of total disability-adjusted life years) in the world. Hemorrhagic stroke (HS) accounts for 37.6% of all stroke types and causes 5.5 million deaths per year approximately, with about half of deaths caused by stroke due to HS. The risk of death from HS is higher compared to ischemic stroke (IS) [3], with a 30-day mortality of 13-61% [4]. In recent years, more and more stroke patients are admitted to the intensive care unit (ICU) for neurological monitoring or

*Correspondence:
Yong Jiang
jy78@vip.sina.com
Baohua Liu
baohualiu@bjmu.edu.cn
[1]Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing, China
[2]Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[3]China National Clinical Research Center for Neurological Diseases, Beijing, China

management of complications, and 10-30% of them are in critical condition [5]. Hence, it is of great significance to optimize the allocation of medical resources by identifying and managing high-risk groups.

Predicting the occurrence of adverse outcomes is the prerequisite for risk stratification. Risk scores are helpful tools for prediction. Many investigators have developed diverse disease risk scoring systems. Traditional scoring systems commonly used in clinical practice include acute physiology and chronic health evaluation(APACHE II) [6], sequential organ failure assessment(SOFA) [7], Oxford acute severity of illness score(OASIS) [8], and simplified acute physiology score(SAPSII) [9], which include various variables with their respective point assignment scheme [10]. However, these traditional scores are applicable to a wide population, whose effectiveness in predicting specific diseases' prognosis is not always satisfactory [11, 12], the application of these scores in HS is limited. Many scholars have made efforts to construct predictive tools for HS. Ho and Smith et al. [13, 14] built a prediction model of HS death in the ICU by logistic regression, and stratified the risk degree of patients by calculating risk scores. However, with the increasing number of clinical examinations and diagnostic items, clinical data often present multidimensional, highly correlated, and nonlinear characteristics [15], which limits the application conditions of traditional clinical modeling methods such as logistic and Cox regression [16]. To compensate for the shortcomings of traditional analytical methods, machine learning algorithms have emerged in the era of big data [17]. Lin and Trevisi et al. [18, 19] employed common machine learning algorithms, such as support vector machine, random forest, and neural network to predict poor functional outcomes in HS patients in the hospital. Howerer, those studies only considered the probability of survival without incorporating the time dimension, by which model prediction is often imprecise [19, 20]. Random survival forest (RSF) is a derivative of the random forest algorithm in survival analysis, which can not only handle complex right-censored survival data but also analyze interactions between variables, and has been applied to pancreatic cancer [21], sepsis [22], and breast cancer [23], and its predictive performance is better than or equivalent to Cox regression in previous studies [22, 24–26], but its performance in the field of HS remains to be investigated.

Therefore, the aim of this study was to establish RSF and Cox regression models based on clinical survival data of HS patients admitted to ICU respectively, and then we evaluated the predictive effect in terms of discrimination and calibration. We also compared developed models with traditional scores with the aim of providing a reference for clinical prediction model construction and clinical decision-making.

Materials and methods

Data source and study participants

All data were extracted from Medical Information Mart for Intensive Care IV, v2.1 (MIMIC-IV) [27], an openly accessible critical care database, containing 256,878 patients electronic medical records, which was collected by Beth Israel Deaconess Medical Center in Boston, Massachusetts from 2008 to 2019. Patients' demographic information, laboratory tests, vital signs, hospital status, medication and surgical procedures were documented in detail in the MIMIC-IV database. All information regarding patients' identification were anonymized and all identifiable information were hidden. Thus, informed consent was exempt. The author had completed all the data research training from the Collaborative Institutional Training Initiative in order to obtain database permission (Record ID:52,310,626). Our study complies with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guideline statement [28].

Among 69,639 ICU admission records in MIMIC-IV, 53,569 patients with first admission were selected. Included criteria were: (1) patients first diagnosed of HS(the diagnosis international classification of diseases (ICD) codes were shown in Table S1); (2) first ICU admission record. Excluded criteria were: (1) patients' age≤18; (2) length of ICU stay≤24 h; (3) patients diagnosed of tumor, cancer or aids. Finally, 2,990 eligible patients were seected for the study in total.

Data collection and outcomes

Based on previous studies [1, 3, 5, 19, 29, 30] and the characteristics of MIMIC-IV database, the extracted variables could be summarized into the following five parts: (1) demographic characteristics: insurance, marital status, admission age, weight, gender, hospital length of stay (HOSLOS), mechanical ventilation, weight; 2) comorbidities: myocardial infarction (MI), congestive heart failure (CHF), peripheral vascular disease (PVD), dementia, chronic pulmonary disease (CPD), rheumatic disease, peptic ulcer disease, mild liver disease, diabetes without chronic complication (diabetes without cc), diabetes with chronic complication (diabetes with cc), paraplegia, renal disease, severe liver disease; 3) vital signs: heart rate, diastolic blood pressure (DBP), systolic blood pressure (SBP), mean blood pressure (MBP), respiratory rate (RR), temperature, peripheral capillary oxygen saturation (Spo2); 4) laboratory examination: hematocrit, hemoglobin, platelets, white blood cell (WBC), anion gap, bicarbonate, blood urea nitrogen (BUN), calcium, chloride, creatinine, glucose, sodium, international normalized ratio (INR), Prothrombin time (PT), partial thromboplastin time (PTT), urine output; 5) conventional scoring systems: SAPSII, OASIS, SIRS, SOFA, Glasgow Coma

Scale (GCS), Charlson comorbidity index (CCI). All of the above variables were recorded within 24 h after ICU admission. If repeated measurements were recorded, the mean values were used instead. For the missing values, if the missing proportion was greater than 20%, the variables will be eliminated, otherwise, the multiple imputation method would be used to fill [31].

Two primary outcomes were analysed in the study. One was the in-hospital survival status within 7 days after admission to the ICU, and the other was the in-hospital survival status within 28 days.

Data preprocessing and statistical analysis

All data were extracted by writing structured query language (SQL) in Navicat Premium 15 software and data analysis procedures were conducted in R 4.1.2 software (R Foundation for Statistical Computing, Vienna, Austria). For continuous variables, normal distribution variables were presented as mean±standard deviation ($\overline{\chi} \pm S$) and t-tests were used for comparisons between groups, while non-normal distribution variables were presented as median and interquartile range [$M(P_{25}, P_{75})$] and Wilcoxon rank-sum test were used for comparison between groups. Categorical variables were presented as frequency and percentages and Chi-square test or Fisher's exact test were used for comparison. A two-side p-value<0.05 was considered statistically significant.

Model construction and evaluation
RSF model

RSF is a non-parametric and nonlinear ensemble learning method, which is the extension of random forest method in survival data [32]. It is an adaptive process that can simulate the complex interaction between nonlinear effects and variables, and find important variables based on the variables ranking of the model's output to reduce generalization errors [33] which makes it well adapted to complex survival data. RSF can calculate the cumulative risk function of each sample even though the assumption of proportional risk is not satisfied, and then aggregate by survival time to generate prediction results of integrated mortality [34].

The procedures to construct a RSF model are as follows: (1) Samples are randomly selected with replacement in the original data by the bootstrap method. Each sample includes 63% of the original data and 37% of the out of bag (OOB) data, and OOB is used to calculate the prediction error rate of the ensemble cumulative risk function. (2) Each bootstrapped sample grows into a survival tree. At each node of the tree, the log-rank or logrank-score splitting criteria is used to select 'mtry' candidate variables which make maximum survival difference between child nodes. When the number of final node events is less than 'nodesize', the growth of survival

tree will quit. (3) The Nelson-Aalen method is used to calculate the cumulative risk for each tree, and the average value of the cumulative risks for all trees is used to estimate the total cumulative risk of the RSF model [34].

Cox regression model

Cox regression is a traditional survival analysis method, which can be presented as $ln\frac{h(t,x)}{h_0(t)} = \beta_1 x_1 + ... + \beta_p x_p$, $h(t, x)$ refers to the risk function at time *'t'* under the influence of risk factor *'t'*, $h_0(t)$ is the baseline risk function at time *'t'* when all independent variables *'x_p'* equal to 0, which is related to time. The mortality risk of each patient is proportional, and the proportional coefficient can be expressed as: $\frac{h(t,x)}{h_0(t)} = exp(\beta_1 x_1 + ... + \beta_p x_p)$. The fitting Cox regression model is usually expressed as: $h(t, x) = h_0(t) exp(b_1 x_1 + ... + b_p x_p)$.
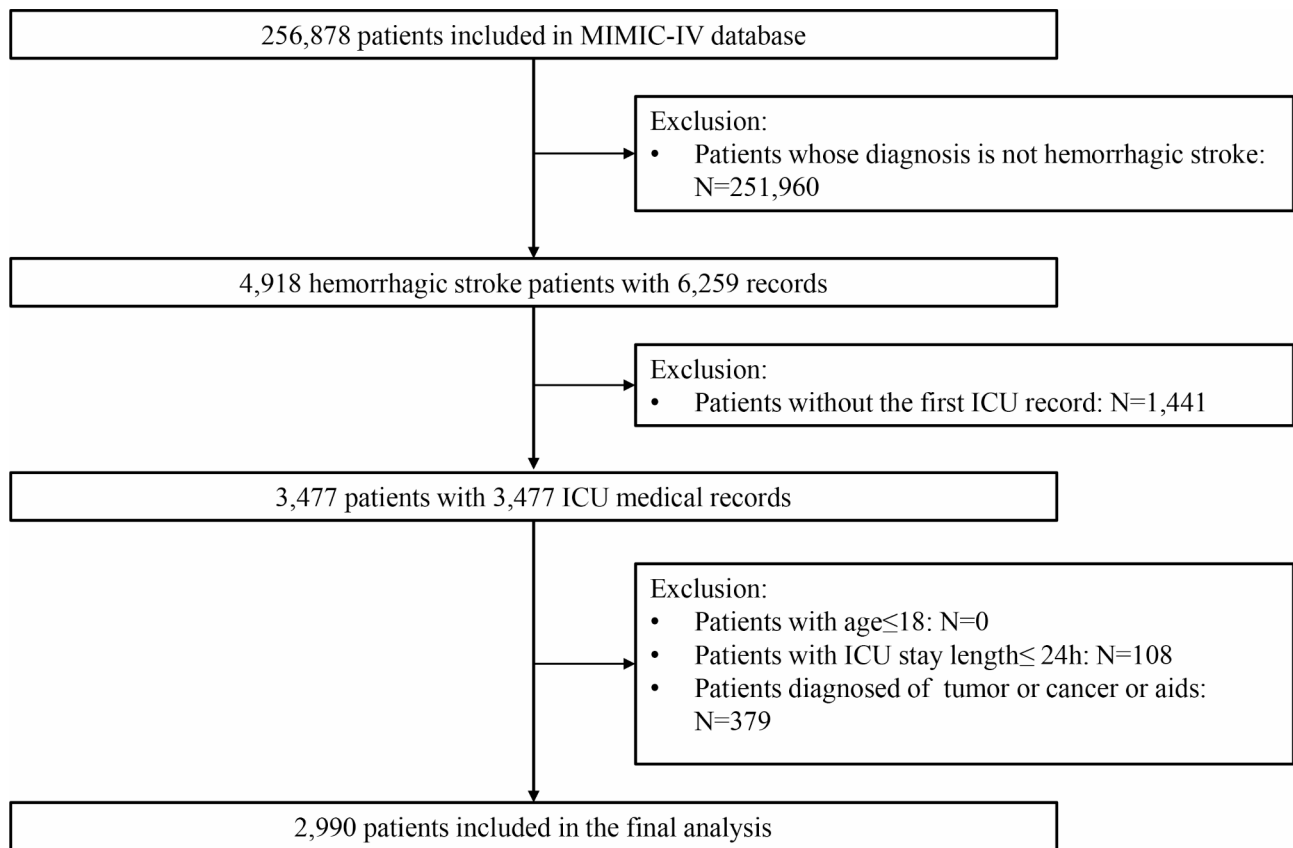
Model construction

All data were randomly divided into the training set and the test set in a ratio of 7:3. In the training set, the least absolute shrinkage and selection operator (LASSO) analysis with ten-fold cross-validation was used to screen variables preliminarily, and then five-fold cross-validation was adopted. In each fold, the grid search method [35] was used to select optimal hyperparameters and the minimum depth method [36] was used to select the optimal variable subset for the construction of RSF model, while the akaike information criterion (AIC) [37] was used to choose the optimal variable subset for Cox regression. The AUC was regarded as the model evaluation index, when the AUC reached optimal value, then the final prediction model was determined.

Model evaluation

In the test set, the bootstrap method [38] with 500 re-samplings was used to compare the differentiation and calibration of RSF and Cox models. The differentiation was evaluated by AUC, and the calibration was evaluated by Brier Score (BS). The larger the AUC was, the better the differentiation was, while, the smaller the BS was, the better the model fit the actual data [39]. Moreover, the positive predictive value (PPV), negative predictive value(NPV) and F1 score were also combined to compare. See Fig. 1 for the model construction and comparison process. In order to further explore the risk stratification ability of Cox and RSF model, patients were divided into low and high risk group according to the optimal cut-off point determined by Yoden index. The Kaplan-Meier (K-M) curve was drawn, and the log-rank test was used to compare whether the survival difference between the two groups was statistically significant. At the same time, the RSF and Cox models were compared with the traditional scores, namely, SAPSII, OASIS, SIRS and SOFA, which further helped to indicate that whether

Fig. 1 Flow chart of model construction and comparison
Notes: LASSO, the least absolute shrinkage and selection operator; HS, hemorrhagic stroke; RSF, Random survival forest; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value

the constructed models had better prediction efficiency. The point assignment scheme of each system were shown in Table S2.

Results

Patient characteristics

A total of 2,990 patients diagnosed with HS were included in our study after excluding those who did not meet the selection criteria, as shown in Fig. 2. Among these patients, 601(20.1%) deaths occurred in hospital after admission to the ICU, 376(12.6%) died within 7 days, and 586(19.6%) died within 28 days. The basic information comparison of these patients stratified by hospital-outcome was shown in Table S3. Compared with those patients who were alive, the dead patients were older($p<0.001$), divorced or widowed($p<0.001$), more likely to be diagnosed with CHF($p<0.001$), liver disease($p<0.001$), diabetes($p=0.037$) and renal disease($p<0.001$), and had a shorter length of stay in hospital. After all patients were randomly divided into the

training and test set, all variables were balanced across the two data sets($p>0.05$), as shown in Table S2.

Variables significance

LASSO regression

After LASSO analysis, a total of 19 variables were screened out from 48 variables (Figure S1): marital status, PVD, severe liver disease, diabetes without cc, admission age, heart rate, MBP, temperature, RR, platelets, WBC, anion gap, creatinine, glucose, sodium, potassium, PT, GCS, weight. The correlation analysis of the selected continuous variables showed that the correlation coefficients among all variables were lower than 0.5 (the lighter the color, the less correlated the variables were), indicating that the correlations between variables were low (Figure S2).

RSF model

After the five-fold cross-validation and grid search in the training set, it was found that when the hyperparameter 'ntree' was set to be 500, 'mtry' was set to be 13, and

Fig. 2 Flow chart of participants inclusion and exclusion
Notes: MIMIC-IV, medical information mart for intensive care IV; ICU, intensive care unit

'nodesize' was set to be 8, the AUC of RSF model reached maximum and the error rate was low as well as stable. Finally, a total of 16 variables were selected: GCS, glucose, admission age, creatinine, temperature, anion gap, RR, sodium, MBP, marital status, heart rate, PT, platelets, potassium, weight, WBC. The error rate and variable importance were shown in Fig. 3, and the grid search process was shown in Figure S3.

Cox regression model

In the training set, five-fold cross-validation was adopted, and backward method was used to screen variables according to AIC. Finally, 18 variables were screened out: WBC, MBP, marital status, GCS, potassium, temperature, anion gap, PVD, severe liver disease, platelets, PT, heart rate, weight, sodium, diabetes without cc, creatinine, admission age, glucose, whose significance could be ranked by absolute value of standardized-β. (Table 1)

Model performance

For the model differentiation, the AUCs of RSF and Cox models in predicting hospital-mortality within 7 days were 0.875(95% CI 0.842–0.908) and 0.794(95% CI 0.724–0.865) respectively, and those in predicting hospital-mortality within 28 days were 0.761(95%CI 0.712–0.809) and 0.649(95%CI 0.559–0.739) respectively. The ROC curves of RSF and Cox models in predicting 7-day and 28-day hospital mortality were shown in Fig. 4. For the model calibration, the brier scores of RSF and Cox were 0.083(95%CI 0.071–0.095) and 0.108(95%CI 0.093–0.123) in predicting hospital-mortality within 7 days, and those in predicting hospital-mortality within 28 days were 0.129(95%CI 0.116–0.143) and 0.174(95%CI 0.167–0.181) respectively. The calibration curve were shown in Figure S4. Other model performance indices, namely, PPV, NPV, F1 score, were shown in Figure S5. In addition, every performance index of the two models was statistically different($p < 0.001$).

Risk stratification

All HS patients were divided into high and low risk group according to optimal cut-off risk-score of RSF and Cox models. The optimal cut-off risk-score of RSF model was 22.20, while that of Cox model was 0.883, and as the patient's risk-score increased, so did the risk of hospital death. K-M curves showed both models could divide patients into high and low risk group with significant survival difference($p < 0.001$). (Figure 5)
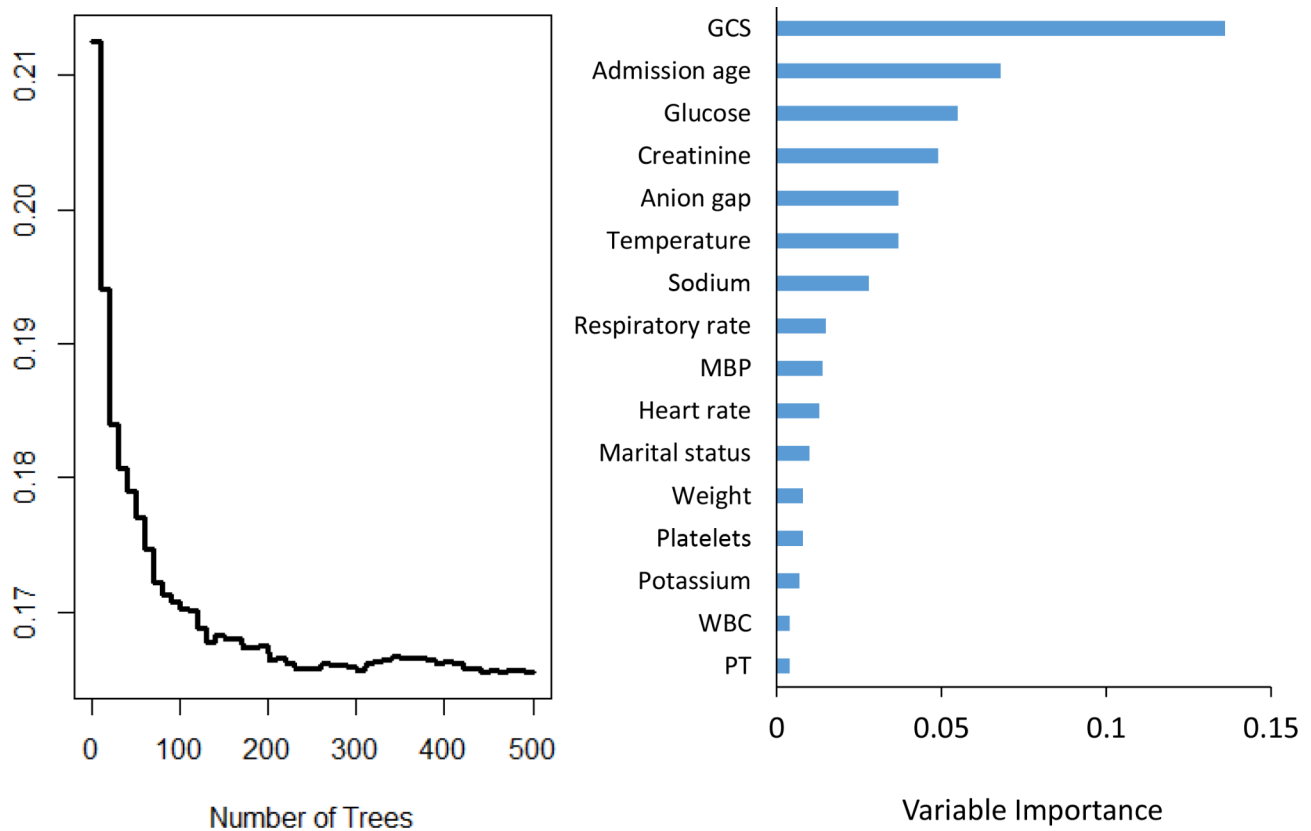
Fig. 3 Error rate curve and variable importance of RSF
Notes: MBP, mean blood pressure; WBC, white blood cell; PT, prothrombin time; GCS, glasgow coma scale

Comparison of RSF with traditional score systems

As the results showed above, the 7-day prediction performance was better than that of 28-day in both RSF and Cox model. Therefore, we further compared the prediction performance of RSF and Cox models with traditional scores, namely, SAPSII, OASIS, SIRS and SOFA. After 500 re-samplings, the box scatter plot showed RSF and Cox models had a good advantage over the four traditional scores in predicting 7-day hospital mortality risk, as shown in Figure S6. The AUCs of six models were 0.875(95%CI 0.842–0.908), 0.761(95%CI 0.712–0.809), 0.736(95%CI 0.685–0.785), 0.723(95%CI 0.676–0.770), 0.632(95%CI 0.575–0.689) and 0.596(95%CI 0.536–0.657) in order from largest to smallest. Other measures related to traditional scoring systems were shown in Table S4. The ROC curves of six models for 7-day mortality were described in Fig. 6. In addition, The AUCs of six models for 28-day mortality were shown in Table S5.

Discussion

In this study, we constructed RSF and Cox regression models to predict the risk of death of HS patients within 7 and 28 days after admission to ICU, respectively. By comparing the AUC, BS, PPV, NPV, and F1 score of the two models, it was found that RSF model was better than

Cox regression model in terms of differentiation and calibration, while both models were able to divide all subjects into low-risk and high-risk groups. Comparing RSF and Cox models with traditional scores, it was observed that the AUC of RSF model was higher than that of traditional scores, suggesting that RSF had a better predictive performance and application value in disease prognosis of HS patients.

For ICU patients, studies have shown that predictive models based solely on survival status do not provide adequate information for clinicians' intervention time decisions [40]. The dependent variables of traditional Cox regression consist of survival time and a dichotomous variable indicating whether survival outcome occurs, which takes into account the patient's survival time but performs poorly when dealing with high-dimension clinical data because the predictive effect is easily limited by the assumption of proportional hazards [37], and can not provide clinicians with a timely decision reference in restricted time with confined medical resources. RSF can be used to generate multiple decision trees by randomizing, integrating all of them to form the final prediction model, which does not depend on applied premises such as *p*-value, proportional hazards assumption, linearity, etc., and reduces computational time by replacing

**Table 1** Multivariate Cox regression analysis results

Characteristics	Standardized β	HR(95% CI)	p value
Marital status(ref:Divorced)			
Married	-0.286	0.751 [0.502, 1.123]	0.164
Single	-0.409	0.664 [0.425, 1.038]	0.073
Widowed	0.005	1.005 [0.636, 1.587]	0.984
PVD(ref:no)	-0.478	0.620 [0.418, 0.920]	0.017
Severe liver disease(ref:no)	0.832	2.299 [1.277, 4.138]	0.006
Diabetes without cc(ref:no)	-0.623	0.536 [0.389, 0.738]	<0.001
Admission age	0.028	1.028 [1.019, 1.038]	<0.001
Heart rate	0.014	1.014 [1.004, 1.023]	0.004
MBP	-0.010	0.990 [0.977, 1.003]	0.124
Temperature	0.239	1.270 [0.999, 1.615]	0.051
Platelets	-0.002	0.998 [0.996, 0.999]	0.011
WBC	0.022	1.022 [0.993, 1.053]	0.143
Anion gap	0.053	1.054 [1.006, 1.104]	0.027
Creatinine	0.310	1.364 [1.196, 1.554]	<0.001
Glucose	0.010	1.010 [1.008, 1.013]	<0.001
Sodium	0.053	1.054 [1.025, 1.085]	<0.001
Potassium	0.249	1.282 [0.994, 1.655]	0.056
PT	0.047	1.048 [1.014, 1.083]	0.005
GCS	-0.027	0.975 [0.947, 1.003]	0.081
Weight	-0.010	0.990 [0.984, 0.997]	0.003

Notes: HR, hazard ratio; PVD, peripheral vascular disease; MBP, mean blood pressure; WBC, white blood cell; PT, prothrombin time; GCS, glasgow coma scale

cross-validation with out-of-bag data estimation. In this study, RSF had higher predictive efficacy than Cox regression for both 7-day and 28-day mortality in HS patients in the ICU. Hadanny et al. predicted 1-year mortality in patients with acute coronary syndrome, and the results also showed that compared to DeepSurv and Cox regression, the RSF model performed best (C-index=0.924) [41], but Qiu et al. obtained the opposite conclusion (C-index=0.611 vs. 0.629) using RSF and Cox regression respectively based on 10 characteristic variables in 82 patients with glioma [42], which may be due to the small sample size and fewer variables in that study. Lin et al. constructed RSF, DeepSurv and Cox to predict the long-term mortality of patients with ischemic stroke, the results showed that all algorithms achieved good prediction effect, which may be because the variables included were few, and these variables satisfied the application conditions of Cox regression. In this case, Cox regression is more suitable for clinical workers to apply and understand [43]. Specially, we found that RSF showed better predictive efficacy for short-term outcomes. Huang et al. also obtained similar findings when using RSF and Cox regression to predict mortality in patients with pot-belly adenocarcinoma over 1–10 years [44], which may be due to the fact that short-term mortality risk is easier to predict because it only needs to consider serious events that occur in a shorter period of time, whereas long-term

mortality risk is subject to more unknown and confounding factors. In addition, compared to traditional scores, we found that the RSF model had the best AUC performance (AUC=0.875), followed by SAPSII, OASIS, SIRS and SOFA. Zhang et al. compared RSF with SOFA, SAPS II and APS III when exploring the 30-day risk of death in sepsis patients in the ICU, and the C-index of the four scores were 0.551, 0.654, 0.669 and 0.731, all of which were lower than the RSF model [22], which was consistent with the results of our study, suggesting that the RSF may provide new methods and ideas for developing new clinical disease severity scores. Moreover, some of the scoring systems had low AUCs, and some others had the opposite in the study, which may due to the fact that SAPSII and OASIS are mainly used to assess the severity of illness and predict prognosis for the ICU or general ward patients. SIRS and SOFA are mainly used for assessing systemic inflammatory response and degree of organ dysfunction respectively.

Identifying variables that are associated with the risk of death has important implications for clinical practice. Both RSF and Cox regression were able to screen for variables of high importance, and among the top 10 important variables, the variables that were identical in both models were creatine, temperature, anon gap, and sodium. Creatine is an important indicator for monitoring of acute kidney injury [45], and Luo et al. showed a steep linear relationship between reduced blood creatine levels and increased risk of in-hospital and 1-year mortality in patients with intracranial hemorrhage when blood creatine values were <1.9 mg/dL [46]. We found a 1.3-fold increase in the risk of in-hospital death in HS patients for each range of temperature change, which was consistent with previous studies [47]. Iglesias Rey et al. conducted a retrospective study of 887 patients with non-traumatic cerebral hemorrhage and found that patients with hypertensive cerebral hemorrhage had the highest body temperature and the greatest increase in body temperature within 24 h. Patients with hypertensive cerebral hemorrhage who developed hyperthermia after 3 months had a 5.3-fold increased risk of poor prognosis, moreover, the amount of edema within 24 h was positively correlated with body temperature in patients with cerebral hemorrhage due to hypertension [48]. Anion gap reflects the acid-base balance in body fluids and plays an important role in the identification of metabolic acidosis [49]. Previous studies have shown that anion gap is an important short- and long-term prognostic marker in patients with IS [50], however, its use in patients with HS is less studied. Shen et al. found that HS patients experienced a decrease in the mini-mental state examination, GCS and other indicators of neurological and cognitive function as the anion gap increased at the time of admission [51]. A meta-analysis had shown that high sodium
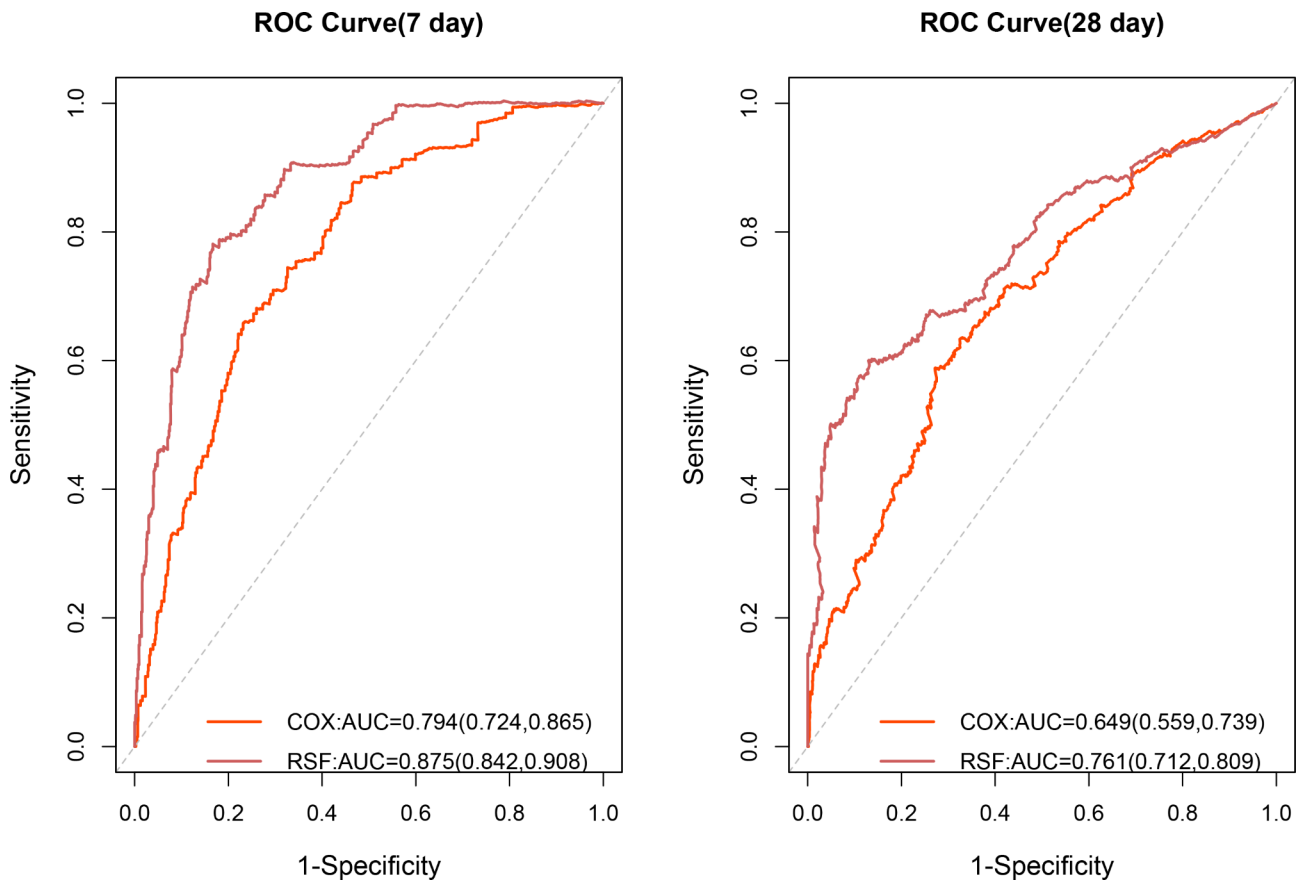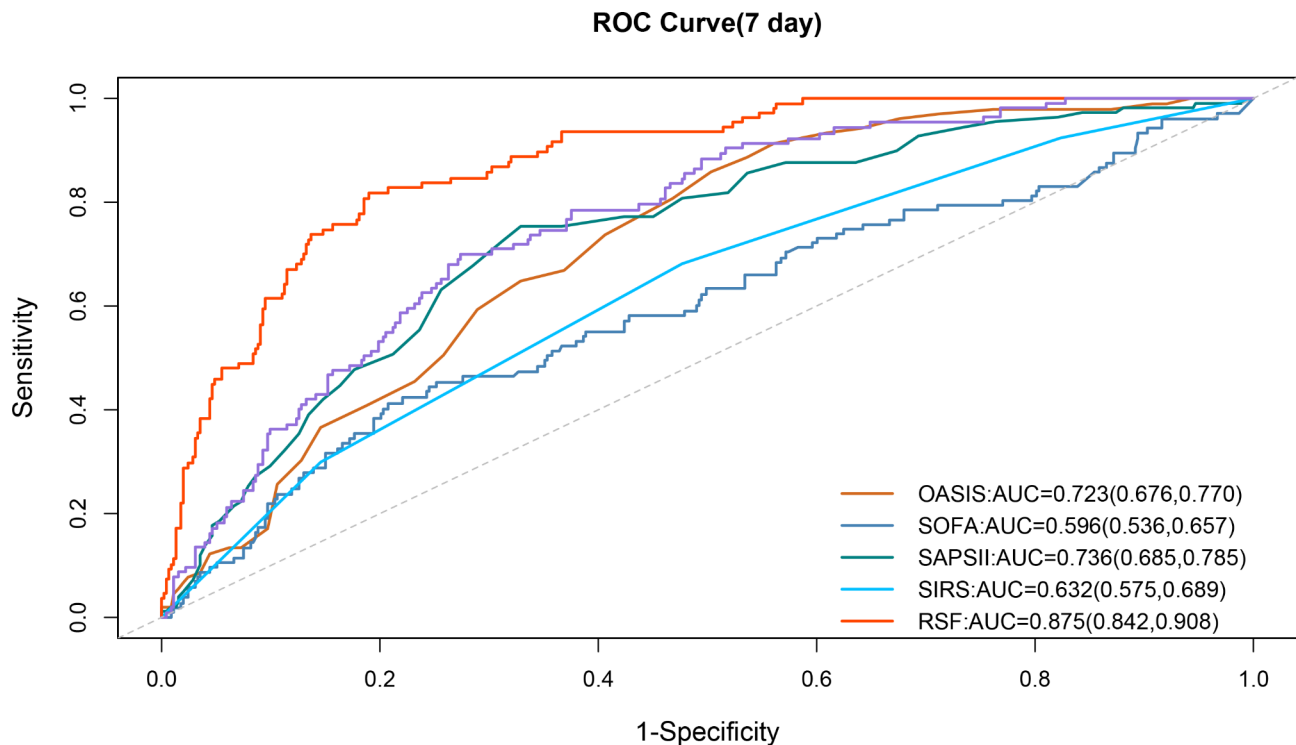
Fig. 4 ROC curve for predicting 7-day and 28-day hospital mortality
Notes: ROC, receiver operation characteristic. RSF, random survival forest
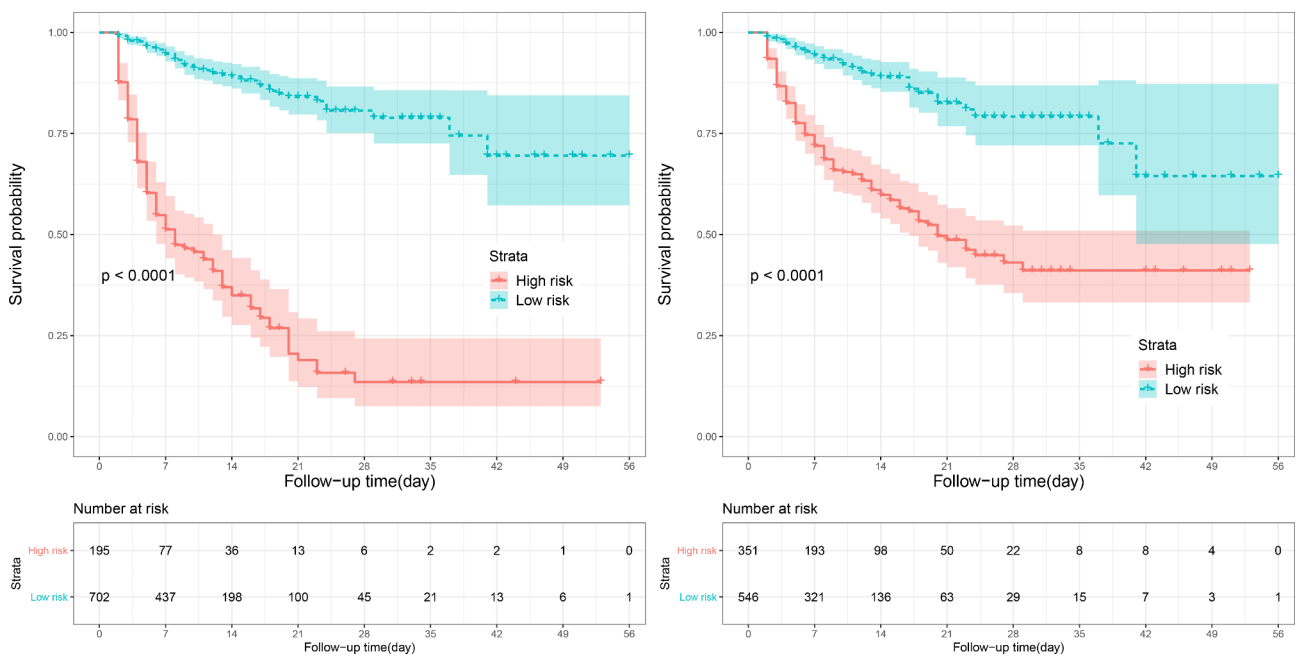
Fig. 5 K-M curves for risk stratification. (**A**) RSF model. (**B**) Cox regression model

ROC Curve(7 day)

Fig. 6 ROC curves of six models

Notes: AUC, area under the curve; RSF, random survival forest; SAPSII, simplified acute physiology score; OASIS, Oxford acute severity of illness score; SIRS, systemic infammactery response syndrome score; SOFA, sequential organ failure assessment. ROC, receiver operation characteristic

intake was positively associated with stroke risk, with a 23% increase in stroke risk for every 86 mmol/d increase in sodium intake [52]. Wang et al., who included 64,909 patients with non-traumatic HS in the United States, showed that spontaneous cerebral hemorrhage patients with abnormal serum sodium had a 1.11-fold increased risk of 30-day readmission compared to patients with normal serum sodium [30].

We have some strengths in this study. We not only compared the predictive efficacy of RSF and Cox, we also compared the models constructed by RSF and Cox with the clinical traditional scoring systems, in addition, we also found the variables that had a strong influence on the occurrence of patient's deaths in the ICU and ranked the variables in terms of importance, which may provide guidance for further practical applications. However, this study has several limitations. Firstly, the MIMIC-IV database is a single-center database, which may limit the applicability of the study results to patients in other centers, so future inclusion of clinical data from multiple centers is desired for external validation. Secondly, due to the limitations of the MIMIC-IV database features, some important indicators such as bilirubin, lactate and albumin could not be included in the analysis because of serious missing values. Finally, only demographic information, laboratory indicators, and comorbidity information were included in this study, and some important

information such as medication and imaging tests were not included, which reduced the predictive performance of the models.

Conclusion

We constructed the RSF and Cox models based on the survival data of patients with HS in the ICU. The results showed that the prediction performance of RSF was better than Cox regression for 7-day and 28-day mortality, with creatine, temperature, anion gap and sodium ranking in the top 10 important variables in both models. RSF can provide new ideas for clinical decision-making of HS patients.

Abbreviations

AUC	Area under the curve
BUN	Blood urea nitrogen
CCI	Charlson comorbidity index
CPD	Chronic pulmonary disease
DBP	Diastolic blood pressure
GCS	Glasgow coma scale
HR	Hazard ratio
HS	Hemorrhagic stroke
INR	International normalized ratio
LASSO	Least absolute shrinkage and selection operator
LOS	Length of stay
MBP	Mean blood pressure
NPV	Negative predictive value
OASIS	Oxford acute severity of illness score
PPV	Positive predictive value
PT	Prothrombin time

PTT	Partial thromboplastin time
PVD	Peripheral vascular disease
ROC	Receiver operation characteristic
RR	Respiratory rate
RSF	Random survival forest
SAPSII	Simplified acute physiology score
SBP	Systolic blood pressure
SIRS	Systemic infammactery response syndrome score
SOFA	Sequential organ failure assessment
Spo2	Peripheral capillary oxygen saturation
WBC	White blood cell

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-023-02293-2.

Supplementary Material 1

Data Availability

MIMIC-IV is a publicly available database. All data in this study can be found at https://physionet.org/about/database/.

Declarations

Ethics approval and consent to participate

The Institutional Review Board of the Beth Israel Deaconess Medical Center and Massachusetts Institute of Technology approved the research use of MIMIC IV for researchers having attended their training course. The author had completed all the data research training from the Collaborative Institutional Training Initiative in order to obtain database permission (Record ID:52310626).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Thayabaranathan T, Kim J, Cadilhac DA, Thrift AG, Donnan GA, Howard G, Howard VJ, Rothwell PM, Feigin V, Norrving B et al. Global stroke statistics 2022. Int J Stroke 2022.
2. Feigin VL, Stark BA, Johnson CO, Roth GA, Bisignano C, Abady GG, Abbasifard M, Abbasi-Kangevari M, Abd-Allah F, Abedi V, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of Disease Study 2019. Lancet Neurol. 2021;20(10):795–820.
3. Akyea RK, Georgiopoulos G, Iyen B, Kai J, Qureshi N, Ntaios G. Comparison of risk of Serious Cardiovascular events after hemorrhagic versus ischemic stroke: a Population-Based study. Thromb Haemost. 2022;122(11):1921–31.
4. van Asch CJ, Luitse MJ, Rinkel GJ, van der Tweel I, Algra A, Klijn CJ. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. Lancet Neurol. 2010;9(2):167–76.
5. Carval T, Garret C, Guillon B, Lascarrou JB, Martin M, Lemarie J, Dupeyrat J, Seguin A, Zambon O, Reignier J, et al. Outcomes of patients admitted to the ICU for acute stroke: a retrospective cohort. Bmc Anesthesiol. 2022;22(1):235.
6. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–29.
7. 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(14):1754–8.
8. Johnson AE, Kramer AA, Clifford GD. A new severity of illness scale using a subset of Acute Physiology and Chronic Health evaluation data elements shows comparable predictive accuracy. Crit Care Med. 2013;41(7):1711–8.
9. 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(24):2957–63.
10. Rahmatinejad Z, Hoseini B, Rahmatinejad F, Abu-Hanna A, Bergquist R, Pourmand A, Miri M, Eslami S. Internal validation of the Predictive performance of Models based on three ED and ICU Scoring Systems to predict Inhospital Mortality for Intensive Care Patients referred from the Emergency Department. Biomed Res Int. 2022;2022:3964063.
11. Afrash MR, Mirbagheri E, Mashoufi M, Kazemi-Arpanahi H. Optimizing prognostic factors of five-year survival in gastric cancer patients using feature selection techniques with machine learning algorithms: a comparative study. BMC Med Inform Decis Mak. 2023;23(1):54.
12. Deng Y, Liu S, Wang Z, Wang Y, Jiang Y, Liu B. Explainable time-series deep learning models for the prediction of mortality, prolonged length of stay and 30-day readmission in intensive care patients. Front Med. 2022;9:933037.
13. Ho WM, Lin JR, Wang HH, Liou CW, Chang KC, Lee JD, Peng TY, Yang JT, Chang YJ, Chang CH, et al. Prediction of in-hospital stroke mortality in critical care unit. SpringerPlus. 2016;5(1):1051.
14. Smith EE, Shobha N, Dai D, Olson DM, Reeves MJ, Saver JL, Hernandez AF, Peterson ED, Fonarow GC, Schwamm LH. A risk score for in-hospital death in patients admitted with ischemic or hemorrhagic stroke. J Am Heart Association. 2013;2(1):e005207.
15. Van Calster B, Wynants L. Machine learning in Medicine. N Engl J Med. 2019;380(26):2588.
16. Steele AJ, Denaxas SC, Shah AD, Hemingway H, Luscombe NM. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. PLoS ONE. 2018;13(8):e0202344.
17. Sabetian G, Azimi A, Kazemi A, Hoseini B, Asmarian N, Khaloo V, Zand F, Masjedi M, Shahriarirad R, Shahriarirad S. Prediction of patients with COVID-19 requiring Intensive Care: a cross-sectional study based on machine-learning Approach from Iran. Indian J Crit care Medicine: peer-reviewed Official Publication Indian Soc Crit Care Med. 2022;26(6):688–95.
18. Lin CH, Hsu KC, Johnson KR, Fann YC, Tsai CH, Sun Y, Lien LM, Chang WL, Chen PL, Lin CL, et al. Evaluation of machine learning methods to stroke outcome prediction using a nationwide disease registry. Comput Methods Programs Biomed. 2020;190:105381.
19. Trevisi G, Caccavella VM, Scerrati A, Signorelli F, Salamone GG, Orsini K, Fasciani C, D'Arrigo S, Auricchio AM, D'Onofrio G, et al. Machine learning model prediction of 6-month functional outcome in elderly patients with intracerebral hemorrhage. Neurosurg Rev. 2022;45(4):2857–67.
20. Jiang L, Zhou L, Yong W, Cui J, Geng W, Chen H, Zou J, Chen Y, Yin X, Chen YC. A deep learning-based model for prediction of hemorrhagic transformation after stroke. Brain Pathol. 2023;33(2):e13023.
21. Lee KS, Jang JY, Yu YD, Heo JS, Han HS, Yoon YS, Kang CM, Hwang HK, Kang S. Usefulness of artificial intelligence for predicting recurrence following surgery for pancreatic cancer: retrospective cohort study. Int J Surg. 2021;93:106050.
22. Zhang L, Huang T, Xu F, Li S, Zheng S, Lyu J, Yin H. Prediction of prognosis in elderly patients with sepsis based on machine learning (random survival forest). BMC Emerg Med. 2022;22(1):26.
23. Xiao J, Mo M, Wang Z, Zhou C, Shen J, Yuan J, He Y, Zheng Y. The application and comparison of machine learning models for the prediction of breast Cancer prognosis: Retrospective Cohort Study. JMIR Med Inform. 2022;10(2):e33440.

24. Bao L, Wang YT, Zhuang JL, Liu AJ, Dong YJ, Chu B, Chen XH, Lu MQ, Shi L, Gao S, et al. Machine learning-based overall survival prediction of Elderly patients with multiple myeloma from Multicentre Real-Life Data. Front Oncol. 2022;12:922039.

25. Grendas LN, Chiapella L, Rodante DE, Daray FM. Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour. J Psychiatr Res. 2021;145:85–91.

26. Pei W, Wang C, Liao H, Chen X, Wei Y, Huang X, Liang X, Bao H, Su D, Jin G. MRI-based random survival forest model improves prediction of progression-free survival to induction chemotherapy plus concurrent Chemoradiotherapy in Locoregionally Advanced nasopharyngeal carcinoma. BMC Cancer. 2022;22(1):739.

27. MIMIC-IV. (version 1.0) [https://mimic.mit.edu/docs/iv/]

28. 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.

29. Carlsson M, Wilsgaard T, Johnsen SH, Johnsen LH, Lochen ML, Njolstad I, Mathiesen EB. Long-term survival, causes of death, and Trends in 5-Year Mortality after Intracerebral Hemorrhage: the Tromso Study. Stroke. 2021;52(12):3883–90.

30. Wang Y, Wang J, Chen S, Li B, Lu X, Li J. Different changing patterns for Stroke Subtype Mortality Attributable to High Sodium Intake in China during 1990 to 2019. Stroke. 2023;54(4):1078–87.

31. Allison PD. Multiple imputation for missing data - A cautionary tale. Sociol Method Res. 2000;28(3):301–9.

32. Yosefian I, Farkhani EM, Baneshi MR. Application of Random Forest Survival Models to increase generalizability of decision trees: a Case Study in Acute myocardial infarction. Comput Math Methods Med. 2015;2015:576413.

33. Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random forests for Survival Analysis using Prediction Error Curves. J Stat Softw. 2012;50(11):1–23.

34. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random Survival Forests. *Wiley StatsRef: Statistics Reference Online* 2019.

35. Wang X, Gong G, Li N, Qiu S. Detection analysis of epileptic EEG using a Novel Random Forest Model Combined with Grid Search optimization. Front Hum Neurosci. 2019;13:52.

36. Ishwaran H. Variable importance in binary regression trees and forests. Electron J Stat. 2007;1:519–37.

37. Kay R. Goodness of fit methods for the proportional hazards regression model: a review. Rev Epidemiol Sante Publique. 1984;32(3–4):185–98.

38. Grunkemeier GL, Wu Y. Bootstrap resampling methods: something for nothing? Ann Thorac Surg. 2004;77(4):1142–4.

39. Murphy AH. A New Decomposition of the Brier score - formulation and interpretation. Mon Weather Rev. 1986;114(12):2671–3.

40. Tang H, Jin Z, Deng J, She Y, Zhong Y, Sun W, Ren Y, Cao N, Chen C. Development and validation of a deep learning model to predict the survival of patients in ICU. J Am Med Inform Assoc. 2022;29(9):1567–76.

41. Hadanny A, Shouval R, Wu J, Gale CP, Unger R, Zahger D, Gottlieb S, Matetzky S, Goldenberg I, Beigel R, et al. Machine learning-based prediction of 1-year mortality for acute coronary syndrome(). J Cardiol. 2022;79(3):342–51.

42. Qiu X, Gao J, Yang J, Hu J, Hu W, Kong L, Lu JJ. A comparison study of machine learning (Random Survival Forest) and Classic Statistic (Cox Proportional Hazards) for Predicting Progression in High-Grade Glioma after Proton and Carbon Ion Radiotherapy. Front Oncol. 2020;10:551420.

43. Lin CH, Kuo YW, Huang YC, Lee M, Huang YW, Kuo CF, Lee JD. Development and validation of a Novel score for Predicting Long-Term Mortality after an Acute Ischemic Stroke. Int J Environ Res Public Health 2023, 20(4).

44. Huang T, Huang L, Yang R, Li S, He N, Feng A, Li L, Lyu J. Machine learning models for predicting survival in patients with ampullary adenocarcinoma. Asia Pac J Oncol Nurs. 2022;9(12):100141.

45. Pickering JW, Frampton CM, Walker RJ, Shaw GM, Endre ZH. Four hour creatinine clearance is better than plasma creatinine for monitoring renal function in critically ill patients. Crit Care. 2012;16(3):R107.

46. Luo H, Yang X, Chen K, Lan S, Liao G, Xu J. Blood creatinine and urea nitrogen at ICU admission and the risk of in-hospital death and 1-year mortality in patients with intracranial hemorrhage. Front Cardiovasc Med. 2022;9:967614.

47. Liddle LJ, Dirks CA, Almekhlafi M, Colbourne F. An ambiguous role for fever in worsening Outcome after Intracerebral Hemorrhage. Transl Stroke Res. 2023;14(2):123–36.

48. Iglesias-Rey R, Rodriguez-Yanez M, Arias S, Santamaria M, Rodriguez-Castro E, Lopez-Dequidt I, Hervella P, Sobrino T, Campos F, Castillo J. Inflammation, edema and poor outcome are associated with hyperthermia in hypertensive intracerebral hemorrhages. Eur J Neurol. 2018;25(9):1161–8.

49. Kraut JA, Nagami GT. The serum anion gap in the evaluation of acid-base disorders: what are its limitations and can its effectiveness be improved? Clin J Am Soc Nephrol. 2013;8(11):2018–24.

50. Liu X, Feng Y, Zhu X, Shi Y, Lin M, Song X, Tu J, Yuan E. Serum anion gap at admission predicts all-cause mortality in critically ill patients with cerebral infarction: evidence from the MIMIC-III database. Biomarkers: Biochem Indic Exposure Response Susceptibility Chemicals. 2020;25(8):725–32.

51. Shen J, Li DL, Yang ZS, Zhang YZ, Li ZY. Anion gap predicts the long-term neurological and cognitive outcomes of spontaneous intracerebral hemorrhage. Eur Rev Med Pharmacol Sci. 2022;26(9):3230–6.

52. Strazzullo P, D'Elia L, Kandala NB, Cappuccio FP. Salt intake, stroke, and cardiovascular disease: meta-analysis of prospective studies. BMJ. 2009;339:b4567.

Publisher's Note