


Comparison and use of explainable machine learning-based survival models for heart failure patients

DIGITAL HEALTH
Volume 10: 1–14
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241277027
journals.sagepub.com/home/dhj



Tao Shi¹, Jianping Yang², Ningli Zhang³, Wei Rong⁴, Lusha Gao¹, Ping Xia¹, Jie Zou¹, Na Zhu¹, Fazhi Yang¹ and Lixing Chen¹ 

Abstract

Objective: Explainable machine learning (XAI) was introduced in this study to improve the interpretability, explainability and transparency of the modelling results. The survex package in R was used to interpret and compare two survival models – the Cox proportional hazards regression (coxph) model and the random survival forest (rfsrc) model – and to estimate overall survival (OS) and its determinants in heart failure (HF) patients using these models.

Methods: We selected 1159 HF patients hospitalised at the First Affiliated Hospital of Kunming Medical University. First, the performance of the two models was investigated using the C-index, the integrated C/D AUC, and the integrated Brier score. Second, a global explanation of the whole cohort was carried out using the time-dependent variable importance and the partial dependence survival profile. Finally, the SurvSHAP(t) and SurvLIME plots and the ceteris paribus survival profile were used to obtain a local explanation for each patient.

Results: By comparing the C-index, the C/D AUC, and the Brier score, this study showed that the model performance of rfsrc was better than coxph. The global explanation of the whole cohort suggests that the C-reactive protein, Ig BNP (brain natriuretic peptide), estimated glomerular filtration rate, albumin, age and blood chloride were significant unfavourable predictors of OS in HF patients in both the coxph and the rfsrc models. By including individual patients in the model, we can provide a local explanation for each patient, which guides the clinician in individualising the patient's treatment.

Conclusion: By comparison, we conclude that the model performance of rfsrc is better than that of coxph. These two predictive models, which address not only the whole population but also selected patients, can help clinicians personalise the treatment of each HF patient according to his or her specific situation.

Keywords

Heart failure, explainable machine learning, survival analysis, Cox proportional hazards regression, random survival forest.

Submission date: 21 March 2024; Acceptance date: 5 August 2024

Introduction

Heart failure (HF) is one of the major pandemics of the twenty-first century and affects more than 64 million people worldwide.¹ With the rapid growth of China's ageing population, the incidence, prevalence and mortality of HF will continue to increase.² Regardless of the cause of HF, it carries enormous social, clinical and economic burdens. Therefore, increasing patient survival and improving patient prognosis are the goals of treating HF.

¹Department of Cardiology, The First Affiliated Hospital of Kunming Medical University, Kunming, China

²College of Big Data, Yunnan Agricultural University, Kunming, China

³Department of Anesthesiology, The Second Affiliated Hospital of Kunming Medical University, Kunming, China

⁴Department of Neurology, The Second Affiliated Hospital of Kunming Medical University, Kunming, China

TS and JY contributed equally as first authors.

Corresponding author:

Lixing Chen, Department of Cardiology, The First Affiliated Hospital of Kunming Medical University, Kunming 650032, China.

Email: ydyyclx@163.com



With the rise of machine learning (ML) technology, an increasing number of clinicians are turning their attention to ML. ML has become a promising tool in health care to help clinicians detect disease patterns, predict patient risk profiles and extract clinical knowledge from massive amounts of data.³ Computer-aided diagnostic systems can accurately and efficiently diagnose complex health problems through the implementation of ML algorithms. However, in addition to specific software and algorithms, survival ML modelling is also constrained by the essential explanation of the model's output. In recent years, many explainable machine learning (XAI) algorithms have been proposed,^{4,5} and in an effort to improve the interpretability, explainability and transparency of ML modelling results, 'black-box' dark ML predictions are transformed into easier-to-interpret 'white-box' glass predictions by XAI algorithms. In this way, physicians can more easily translate complex information in ML models into reliable daily clinical choices.^{6,7}

Recently, the XAI method has been applied to survival models, and researchers have developed several R packages for survival models.^{8–10} Of these, the *survex* package is a recent development that has not been applied to the prediction of survival in HF patients. Therefore, the aim of this study was to use the *survex* package to interpret and compare two survival models, the Cox proportional hazards regression model (*coxph*) and the random survival forest (*rfsrc*) model,^{11,12} and to use this approach to estimate overall survival (OS) and its determinants in HF patients.

Materials and methods

Study population

As a rough estimate, the sample size should be 10 times the number of independent variables, and this study involves 41 independent variables, so the sample size should be greater than 410. This study is retrospective. Our study included 1221 HF patients who were hospitalised at the First Affiliated Hospital of Kunming Medical University between January 2017 and October 2021 due to acute exacerbation of chronic HF. These patients had New York Heart Association (NYHA) functional class III or IV disease. The following patients were excluded: (i) patients who lacked the required data; (ii) patients who had severe comorbidities (malignancy, infectious, haematological, or severe hepatic or renal dysfunction) and (iii) patients who were lost to follow up. In the end, 1159 patients were included in this study.

Data collection

At the time of admission, we collected basic information, history, current treatment, laboratory results, electrocardiogram data and cardiac ultrasound data from the patients. A total of 104 variables were collected for each patient. To

determine the more holistic clinical characteristics of the study population, we identified 41 variables potentially associated with HF to analyse the baseline population characteristics; these variables have been frequently used in other studies.^{13–15} Ultimately, we selected 8 additional important variables from these 41 variables that have been shown to affect survival and prognosis in HF patients and used them to create two additional survival models (*coxph* and *rfsrc*). In a study by Sciomer et al.,¹⁶ age was shown to affect the prognosis of HF patients. A low albumin concentration is associated with a poorer prognosis in HF patients.¹³ CRP (C-reactive protein) and BNP (brain natriuretic peptide) are independent risk factors in HF patients.^{17,18} Low serum sodium and low serum chloride increase mortality in HF patients.^{19,20} A study by Tedeschi et al.²¹ demonstrated that chronic kidney disease and elevated uric acid worsened the prognosis of HF patients. In the end, we chose 8 variables: age, serum albumin concentration, CRP concentration, Ig BNP concentration, blood sodium concentration, blood chloride concentration, uric acid concentration and eGFR (estimated glomerular filtration rate). All blood samples were collected and delivered to the laboratory of the First Affiliated Hospital of Kunming Medical University after the patients had fasted overnight (8–12 hours).

Through phone conversations with patients or their relatives, the investigators gathered survival data; those who did not answer the phone were considered lost. Written informed consent was obtained from all patients before the start of this study.

Algorithm development

The traditional machine learning models are 'black box' models with low interpretability and transparency. To increase the breadth of model explainability tools accessible in the R environment, packages such as DALEX and *iml* have been developed, which provide a comprehensive variety of explainable machine learning approaches. While there are several programmes in this field, their primary focus is on describing classification and regression models. Certain explanatory approaches may be used for survival models, but careful adjustments are required due to the models' unique predictive character. To overcome this issue, *survex* provides precisely customised explanations that reflect the temporal dimension inherent in survival model predictions. Furthermore, methods for explaining survival models have been developed, such as *SurvLIME* with a Python implementation and *SurvSHAP(t)*. The *survex* package is likewise equipped to embrace these new procedures, which expands its functional range. It is worth noting that the *survxai* package served as a key inspiration for the creation of *survex*. *Survex*, on the other hand, provides a greater variety of features, freshly developed explanation approaches, and model support.

Statistical analysis

The primary outcome of this study was OS, which was the time from discharge to death from any cause or to the last follow-up visit for HF patients. We analysed OS using coxph and rfsrc and then used the XAI to explore survival and interpret the results.

For the baseline characteristics of the patients, continuous variables are expressed as the mean \pm standard deviation if normally distributed; otherwise, they are expressed as the median with interquartile range. The categorical variables are presented as numbers and percentages. The distribution of the original BNP values was strongly skewed, so the values were log-transformed. Independent sample t tests were used for normally distributed continuous variables, Mann–Whitney U tests were used for nonnormally distributed data, and Chi-square tests were used for categorical variables. A variable was considered to be statistically significant when its p value was <0.05 .

For both the coxph and rfsrc models, model performance was first examined using the consistency C-index, the integrated cumulative/dynamic AUC and the integrated Brier score. Subsequently, a global explanation of the whole cohort was carried out using the time-dependent variable importance and the partial dependence survival profile. Finally, the local explanation for a single patient was derived from the SurvSHAP(t) and SurvLIME plots and the ceteris paribus survival profile. In all the graphs, the X-axis indicates the period between discharge and any cause of mortality or the final follow-up visit, where all the event times are highlighted in red and the census times are highlighted in grey.

All the statistical analyses for this study were performed with the IBM SPSS Statistics version 26.0, R 4.3.2, survexp, survival, randomForestSRC, and MICE packages.

Results

Study population characteristics

A total of 1159 HF patients were included in this analysis. The median follow-up time for this study was 759 days. Ultimately, 536 patients died and 623 survived. The mean age of the included patients was 66.83 ± 12.52 years; 719 (62.0%) were males and 440 (38.0%) females. Specific differences between the deceased group and the survivor group are described in Table 1.

Model performance for the whole cohort

We constructed two different survival models (coxph and rfsrc) using eight clearly reported factors affecting the prognosis of HF for survival analysis. The eight factors were age, lg BNP, blood sodium, blood chloride, albumin, uric acid, eGFR and CRP. We used Brier scores, C/D AUCs

and C-indices to estimate model performance. The values of all these metrics are shown on the y-axis. Higher C/D AUC and C-index values are associated with better model performance, while lower Brier scores are associated with better model performance. For coxph, the C-index was 0.729, the C/D AUC was 0.708, and the Brier score was 0.173. For rfsrc, the C-index was 0.823, the C/D AUC was 0.818, and the Brier score was 0.120. As shown above, the rfsrc outperformed coxph both for each metric (Figure 1) and throughout the follow up (Figure 2). Compared to coxph, rfsrc was more predictive, but it was also less interpretable. Table 2 shows the results of the univariate and multivariate coxph models consisting of these eight factors. Cox multivariate proportional hazards models showed that all variables except for blood sodium were significantly different.

Global explanation: Time-dependent feature importance for the whole cohort

For the coxph and rfsrc models, we used two different methods to examine the importance of time-dependent variables for the whole cohort. Variable interpretation was performed after splitting each variable. The y-axis indicates the change in the loss function after substitution for each covariate. Variable importance can change over time, with higher values of the loss function indicating a greater influence of the variable on OS. By using the Brier score loss after permutation (Figure 3) and the C/D AUC loss after permutation (Figure 4), we found that in coxph, lg BNP was the worst independent risk factor for OS, while in rfsrc, CRP was the worst independent risk factor for OS.

Global explanation: Partial dependence survival profile for the whole cohort

The partial dependence survival profiles (PDPs) for the coxph and rfsrc models are shown in Figures 5 and 6, respectively, which indicate how the OS of the whole cohort changes if the value of one determinant is altered but all other factors are held constant. The y-axis represents the value of the survival function for each covariate. A wider area of the curve indicates that the greater the difference in levels of a factor, the greater the effect of that factor on OS. CRP, lg BNP, eGFR, albumin, age and blood chloride concentration were found to be very important for OS in both the coxph and rfsrc models, with CRP being the most important unfavourable predictor.

Local explanation: SurvSHAP(t) plot for a single patient

Using SurvSHAP(t) plots, it is possible to investigate the time-dependent survival contribution of each risk factor to

Table 1. Baseline characteristics.

Variables	Total (n = 1159)	Deceased group (n = 536)	Survivor group (n = 623)	p Value
Basic characteristics				
Age (year)	66.83 ± 12.52	69.71 ± 11.96	64.24 ± 12.37	<0.001
Male	719 (62.0%)	334 (62.3%)	385 (61.8%)	0.857
Systolic BP (mmHg)	122.10 ± 22.94	120.19 ± 23.54	123.67 ± 22.00	0.009
Diastolic BP (mmHg)	76.23 ± 15.04	74.11 ± 14.79	77.96 ± 15.85	<0.001
HR (beat/minute)	85.24 ± 21.02	86.73 ± 20.85	83.62 ± 20.72	0.011
BMI (kg/m ²)	23.02 ± 3.81	22.40 ± 3.44	23.56 ± 4.04	<0.001
NYHA class				<0.001
III	734 (63.3%)	263 (49.1%)	471 (75.6%)	
IV	425 (36.7%)	273 (50.9%)	152 (24.4%)	
Medical history				
Coronary disease	592 (51.1%)	275 (51.3%)	317 (50.9%)	0.886
Hypertension	641 (55.3%)	294 (54.9%)	347 (55.7%)	0.772
Diabetes	331 (28.6%)	172 (32.1%)	159 (25.5%)	0.014
Atrial fibrillation	398 (34.3%)	195 (36.4%)	203 (32.6%)	0.175
Stroke	163 (14.1%)	82 (15.3%)	81 (13.0%)	0.262
Laboratory indicators				
WBC (10 ⁹ /L)	6.88 (5.51, 9.05)	7.13 (5.65, 9.83)	6.75 (5.47, 8.60)	0.015
RBC (10 ¹² /L)	4.53 ± 0.77	4.44 ± 0.82	4.63 ± 0.71	<0.001
NBC (10 ⁹ /L)	4.50 (3.47, 6.45)	4.81 (3.49, 7.11)	4.34 (3.46, 5.78)	0.001
LBC (10 ⁹ /L)	1.39 (1.01, 1.83)	1.25 (0.88, 1.70)	1.49 (1.12, 1.90)	<0.001
CRP (mg/L)	7.40 (3.00, 21.36)	15.47 (5.80, 34.97)	4.47 (2.18, 11.80)	<0.001
Hb (g/L)	137.92 ± 21.12	134.24 ± 26.10	141.09 ± 21.82	<0.001
PLT (10 ⁹ /L)	191.00 (147.00, 243.00)	182.00 (132.00, 234.75)	201.00 (159.00, 248.00)	<0.001
Sodium (mmol/L)	141.08 ± 4.40	140.27 ± 4.87	141.77 ± 3.81	<0.001
Potassium (mmol/L)	3.94 ± 0.60	3.98 ± 0.67	3.91 ± 0.54	0.047
Chloride (mmol/L)	102.92 ± 4.67	101.79 ± 4.90	103.89 ± 4.24	<0.001

(continued)

Table 1. Continued.

Variables	Total (n = 1159)	Deceased group (n = 536)	Survivor group (n = 623)	p Value
Alb (g/L)	36.68 ± 4.56	35.77 ± 4.90	37.46 ± 4.09	<0.001
ALT (IU/L)	25.00 (16.50, 42.30)	24.15 (16.00, 42.30)	25.00 (16.80, 42.30)	0.437
AST (IU/L)	28.20 (20.00, 43.00)	28.95 (20.3, 51.15)	27.20 (20.00, 40.00)	0.023
Cre (µmol/L)	103.00 (83.00, 133.10)	109.45 (87.63, 145.38)	97.60 (79.30, 121.80)	<0.001
UA (µmol/L)	477.40 (371.20, 588.20)	509.50 (406.03, 618.00)	457.50 (356.00, 563.30)	<0.001
TC (mmol/L)	3.67 ± 1.00	3.53 ± 1.00	3.80 ± 0.99	<0.001
eGFR (ml/min)	44.26 (32.73, 56.81)	39.96 (27.48, 51.77)	48.26 (37.27, 60.65)	<0.001
Ig BNP (pg/ml)	3.17 ± 0.28	3.24 ± 0.29	3.10 ± 0.25	<0.001
ECG parameters and cardiac ultrasound index				
QRS wave (ms)	115.41 ± 29.81	118.32 ± 32.46	112.91 ± 27.12	0.002
LAd (mm)	42.44 ± 9.38	40.09 ± 9.81	41.89 ± 8.96	0.030
LVDd (mm)	56.29 ± 12.73	56.26 ± 13.32	56.32 ± 12.22	0.935
RAAd (mm)	52.06 ± 12.60	52.68 ± 13.69	51.52 ± 11.58	0.124
RVd (mm)	67.41 ± 16.63	66.44 ± 17.84	68.23 ± 15.49	0.072
LVEF (%)	43.00 (33.00, 58.00)	42.00 (31.00, 56.00)	45.00 (35.00, 59.00)	0.007
Treatment				
SGLT-2I	262 (22.6%)	131 (24.4%)	131 (21.0%)	0.166
β-Receptor blockers	812 (70.1%)	368 (68.7%)	444 (71.3%)	0.333
Diuretics	913 (78.8%)	416 (77.6%)	497 (79.8%)	0.369
ACEI/ARB/ARNI	650 (56.1%)	307 (57.3%)	343 (55.1%)	0.448
CRT/CRTD	113 (9.7%)	66 (12.3%)	47 (7.5%)	0.006

Note: When comparing differences in continuous variables, the independent samples t tests were used if normally distributed, and the Mann-Whitney U tests were used if nonnormally distributed; differences between groups in categorical variables were compared using Chi-squared tests; p values were derived by comparing the deceased group with the survivor group; p value < 0.05 indicated statistical significance.

ACEI, angiotensin converting enzyme inhibitor; Alb, albumin; ALT, alanine aminotransferase; ARB, angiotensin II receptor blocker; ARNI, angiotensin receptor-enkephalinase inhibitor; AST, aspartate aminotransferase; BMI, body mass index; BNP, brain natriuretic peptide; BP, blood pressure; Cre, creatinine; CRP, C-reactive protein; CRT, cardiac resynchronisation therapy; CRTD, cardiac resynchronisation therapy defibrillator; eGFR, estimated glomerular filtration rate; Hb, haemoglobin; HR, heart rate; LAd, left atrium diameter; LBC, lymphocyte; LVDd, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; NBC, neutrophil; NYHA, New York Heart Association; PLT, platelet; RAAd, right atrium diameter; RBC, red blood cells; RVd, right ventricle diameter; SGLT-2I, sodium-glucose cotransporter 2 inhibitor; TC, total cholesterol; UA, uric acid; WBC, white blood cells.

OS for a selected patient. The y-axis represents the SurvSHAP(t) value for each variable, where a positive number indicates that the variable increases the OS of that

patient, and conversely, a negative number indicates that the variable decreases the OS of that patient. We used HF patient #164 (81 years old, Ig BNP 3.18 pg/ml, blood

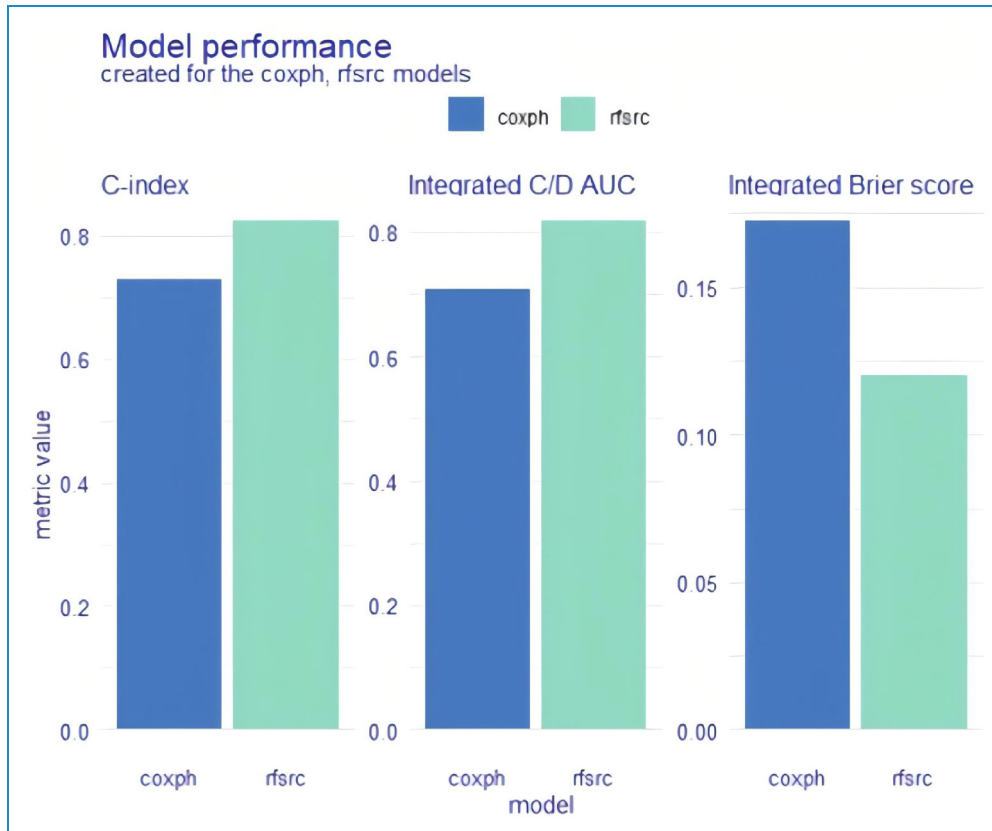


Figure 1. Model performance for the whole cohort. Explainable machine learning (XAI) data are shown as bar plots. coxph: Cox proportional hazards regression; rfsrc: random survival forest.

sodium 137.3 mmol/L, blood chloride 105.7 mmol/L, albumin 36.8 g/L, uric acid 498.7 $\mu\text{mol/L}$, eGFR 19.99 ml/min and CRP 7.8 mg/L) in the survival model to achieve a shift from predictions for the whole cohort to individual patients. The SurvSHAP(t) plot for patient #164 showed that, in coxph, blood chloride improved this patient's odds of survival, and the eGFR decreased this patient's odds of survival; in rfsrc, CRP improved this patient's odds of survival, and Ig BNP decreased this patient's odds of survival (Figure 7).

Local explanation: SurvLIME plot for a single patient

The SurvLIME plot is very similar to the SurvSHAP(t) plot in that both can be used to detect the survival predictors that have the most impact on a single selected patient. There are two parts to the SurvLIME plot. The left part shows the effect of each variable on the survival of a single selected patient: the larger the area, the greater the influence; the higher the value of the local importance of SurvLIME, the lower the chances of survival. The right part shows a comparison between the predictions of the coxph or rfsrc models and the predictions of the black-box model: the closer the two functions, the more accurately the model's

results are explained. We still included patient #164 in the coxph and rfsrc models to obtain two SurvLIME plots (Figures 8 and 9). From Figure 8, we can conclude that, in coxph, albumin increases the survival chances of this patient, and Ig BNP decreases the survival chances of this patient; these two functions gradually separated after 500 days, which indicated that the accuracy of the model's survival estimation for this patient gradually decreased after 500 days. Figure 9 shows that in the rfsrc cohort, blood sodium increases patient survival, and Ig BNP decreases patient survival; these two functions are relatively close to each other, which means that the survival estimation of the patient is relatively accurate.

Local explanation: Ceteris paribus, survival profile for a single patient

The ceteris paribus survival profile (CPP) can be seen as an equivalent of the PDP, but it should be used for a single observation. As with PDP, the lower the y-axis value of the CPP survival function, the worse the OS, and the variables with the largest differences between levels were those that had the greatest impact on OS. Figures 10 and 11 show the CPPs for coxph and rfsrc, respectively, for patient #164,

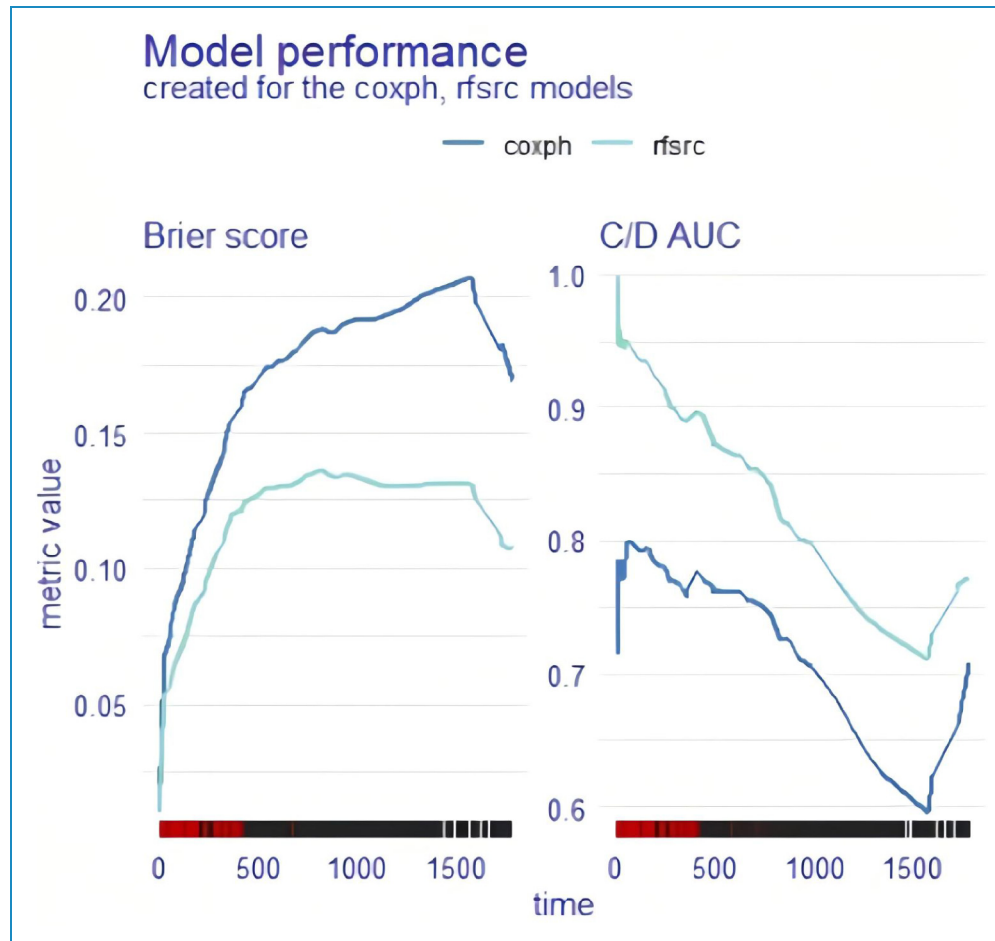


Figure 2. Model performance for the whole cohort. Explainable machine learning (XAI) was used as a time-dependent estimation. coxph: Cox proportional hazards regression; rfsrc: random survival forest.

Table 2. Univariate and multivariate Cox proportional hazards model for overall survival in heart failure patients.

	Univariate		Multivariate	
	HR (95% CI)	p Value	HR (95% CI)	p Value
Age	1.032 (1.024, 1.040)	<0.001	1.024 (1.015, 1.033)	<0.001
Ig BNP	5.517 (4.030, 7.552)	<0.001	3.206 (2.313, 4.444)	<0.001
Blood sodium	0.942 (0.924, 0.961)	<0.001	0.996 (0.973, 1.019)	0.741
Blood chloride	0.929 (0.912, 0.946)	<0.001	0.946 (0.926, 0.967)	<0.001
Albumin	0.928 (0.910, 0.946)	<0.001	0.952 (0.932, 0.972)	<0.001
Uric acid	1.001 (1.001, 1.002)	<0.001	1.001 (1.000, 1.001)	0.006
eGFR	0.976 (0.971, 0.981)	<0.001	0.993 (0.987, 0.999)	0.021
C-reactive protein	1.011 (1.009, 1.012)	<0.001	1.008 (1.006, 1.010)	<0.001

BNP: brain natriuretic peptide; CI: confidence interval; eGFR: estimated glomerular filtration rate; HR: hazard ratio.

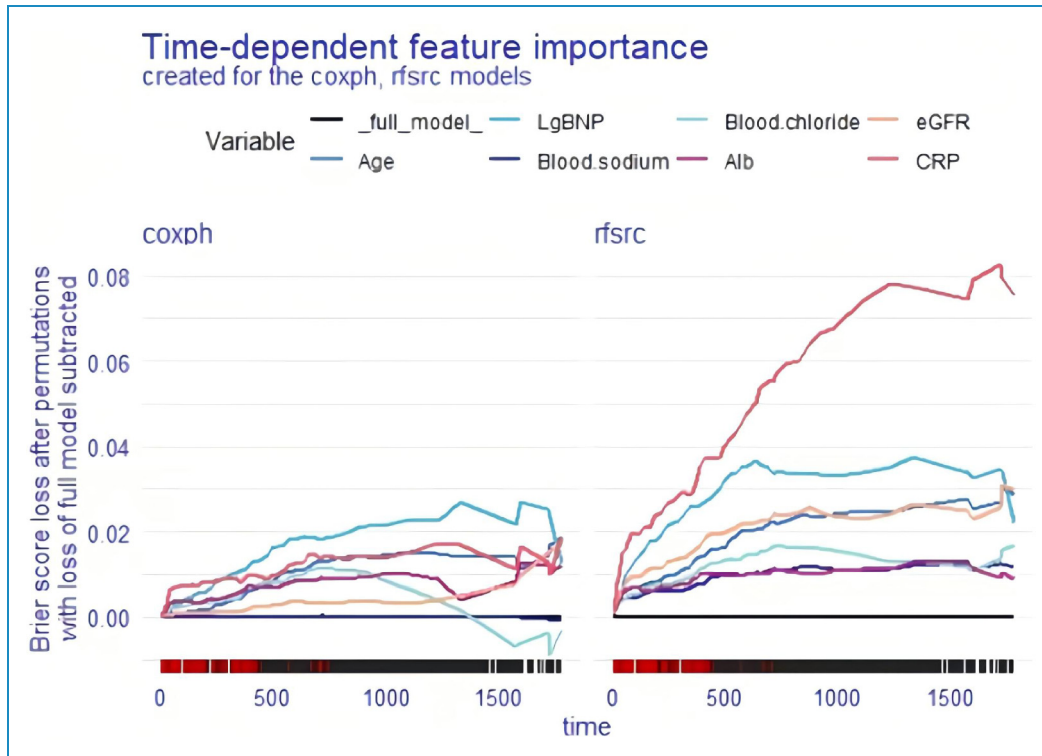


Figure 3. Global explanation: time-dependent feature importance for the whole cohort and Brier score loss after permutation. Alb: albumin; BNP: brain natriuretic peptide; coxph: Cox proportional hazards regression; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; rfsrc: random survival forest.

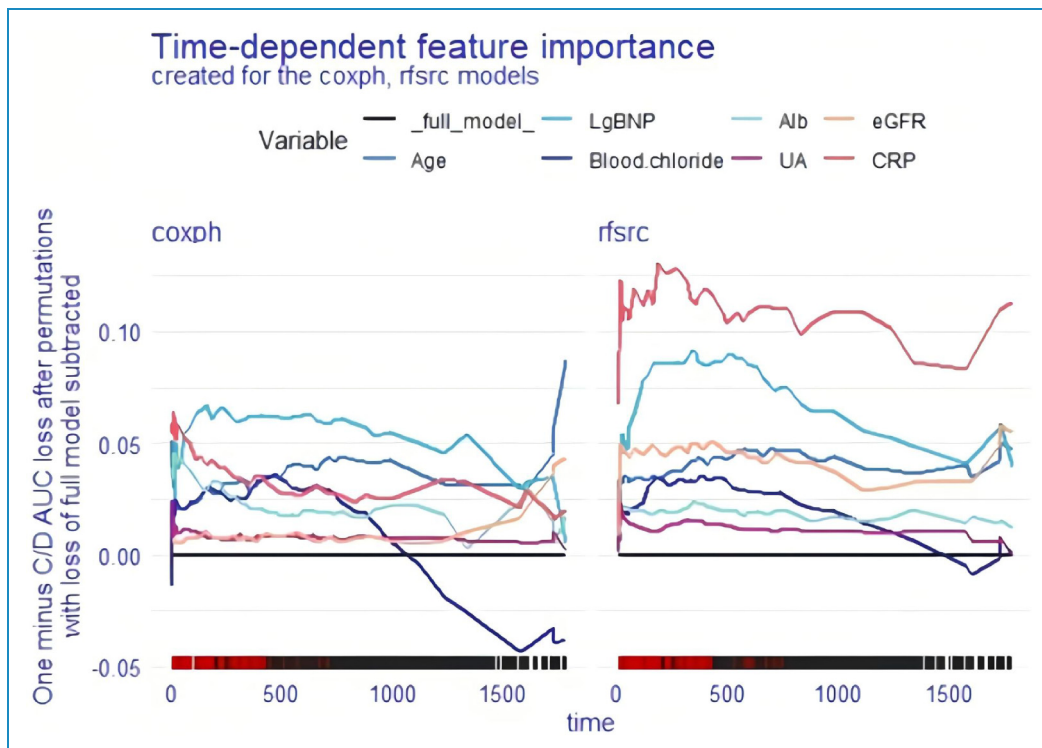


Figure 4. Global explanation: time-dependent feature importance for the whole cohort and C/D AUC loss after permutation. Alb: albumin; BNP: brain natriuretic peptide; coxph: Cox proportional hazards regression; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; rfsrc: random survival forest; UA: uric acid.

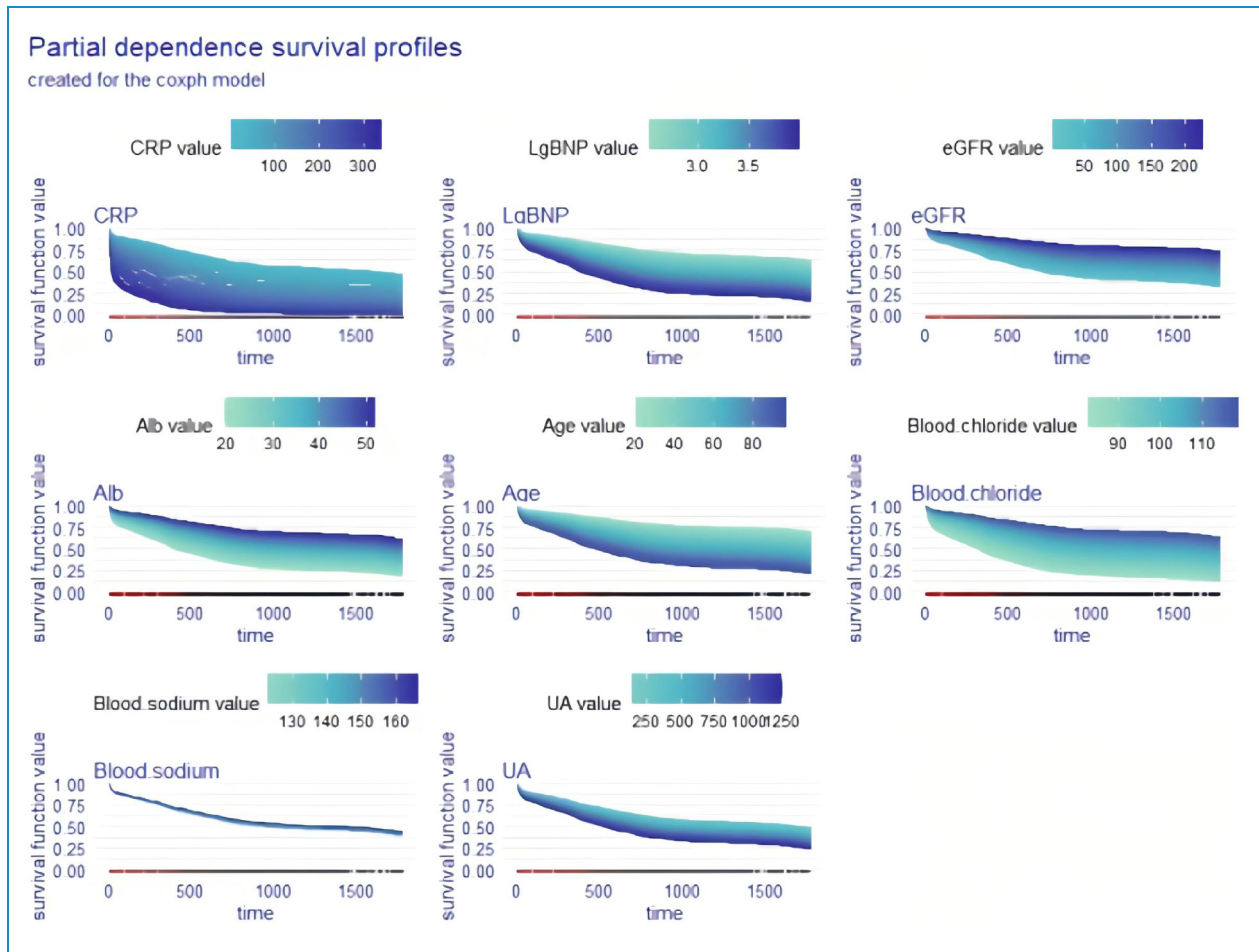


Figure 5. Global explanation: PDP for the whole cohort; coxph model. Alb: albumin; BNP: brain natriuretic peptide; coxph: Cox proportional hazards regression; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; PDP: partial dependence survival profile; UA: uric acid.

with the patient's values in the model shown as red lines. According to the coxph model, all factors except for the serum sodium concentration affected OS. According to the rfsrc model, lg BNP had the greatest effect on OS.

Discussion

The medical field generates a large amount of data, including patient records, images, physiological signals and genetic data. ML can help doctors and researchers extract useful information from complex data and identify potential patterns and associations. ML is also increasingly being used in the field of HF.^{22–24} ML can help clinicians identify potential associations between HF survival and clinical indicators, allowing for the early identification of high-risk patients and the provision of appropriate treatment options. While the integration of ML into biomedical research and health care holds great promise, the opaque nature of black-box models has caused reasonable concerns. To solve this problem, researchers have developed several XAI methods. Among them, the survex package in R

developed by Mikolaj Spytek's team is a very useful tool that provides a large and complete selection of ML model interpreters.⁹ With the survex software package, all analyses may be performed under a unified interface, which enables the exploration and easy comparison of the outcomes derived from different survival models (coxph and rfsrc) for HF patients.

First, we examined the model quality of coxph and rfsrc by measuring their performance with three different metrics: the C-index, the integrated C/D AUC, and the integrated Brier score. Our study showed that, compared with coxph, rfsrc had higher C-index and C/D AUC values and lower Brier score. Thus, the performance of the rfsrc model is better and more predictive than that of the other models, but the model is also less interpretable.

Second, a range of global explanations of the coxph and rfsrc models were carried out to explore the predictions of the models for the whole patient cohort. Two different loss functions, the Brier score and 1-CD/AUC, were used to estimate the importance of each mutual variable, involving a time-dependent process. The Brier score loss after

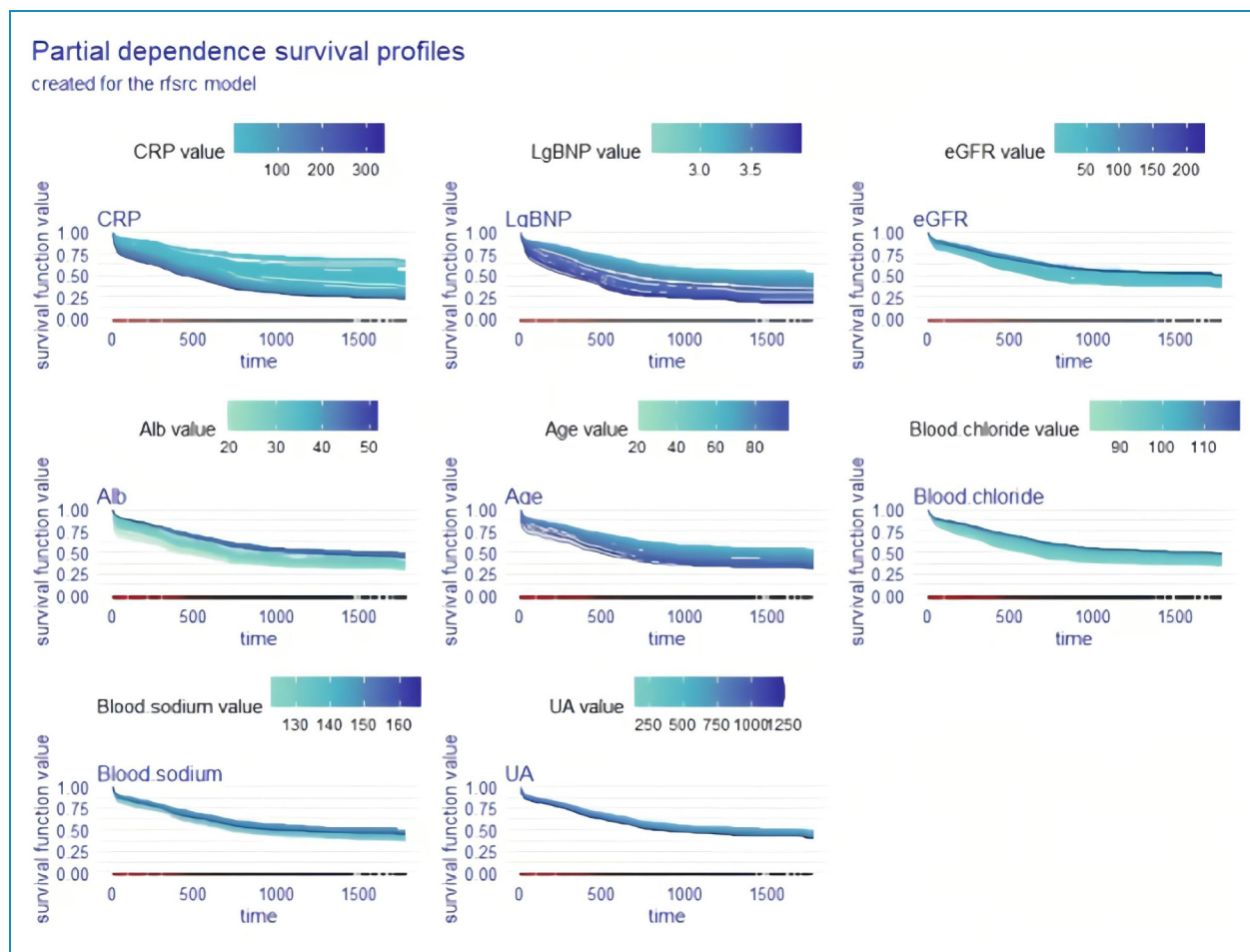


Figure 6. Global explanation: PDP for the whole cohort; rfsrc model. Alb: albumin; BNP: brain natriuretic peptide; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; PDP: partial dependence survival profile; rfsrc: random survival forest; UA: uric acid.

permutation and the C/D AUC loss after permutation showed that lg BNP was the worst independent risk factor for OS in coxph-modelled patients and that CRP was the worst independent risk factor for OS in rfsrc-modelled patients. The PDPs showed that CRP, lg BNP, eGFR, albumin, age and blood chloride concentration were important adverse predictors of OS in HF patients, whether in the coxph or rfsrc model.

Finally, the model's predictions for a single selected observation can be explored through another series of local explanations. The SurvSHAP(t) is based on SHAP values, an additivity interpretation method based on Shapley values from cooperative game theory.²⁵ The SurvSHAP(t) allows the analysis of the impact of variables on model predictions at different points in time, providing a unique perspective on understanding survival function predictions. The SurvSHAP(t) quantifies the importance of each feature in the model by calculating its contribution to the prediction, making the predictions more interpretable.²⁶ The SurvLIME evaluates local interpretability by evaluating attributes for a specific individual. It differs

from prior techniques in that it takes into account temporal space for explanation. It starts by creating a set of neighbours, gets a set of predictions for the neighbours, and then fits a coxph to minimise the difference between the black box model's predictions and the local explainer's predictions.²⁷ The SurvSHAP(t) function may show the effect of each risk factor on OS for the selected patient. The function of SurvLIME allows us to understand the importance of each factor for OS and its favourable or unfavourable impact on the time course. With the CPP, it is possible to graphically estimate the effect of each risk factor on the OS of any individual observation.

The three steps above enabled us to determine which factors influence the OS of HF patients and the extent of their influence. We can also find the most important risk factors affecting OS for a single selected patient so that a personalised treatment plan can be developed for the patient.

There are several limitations to our study. First, the use of retrospective data for model construction in the study raises the possibility that some potential risks were not taken into account; thus, further prospective experiments

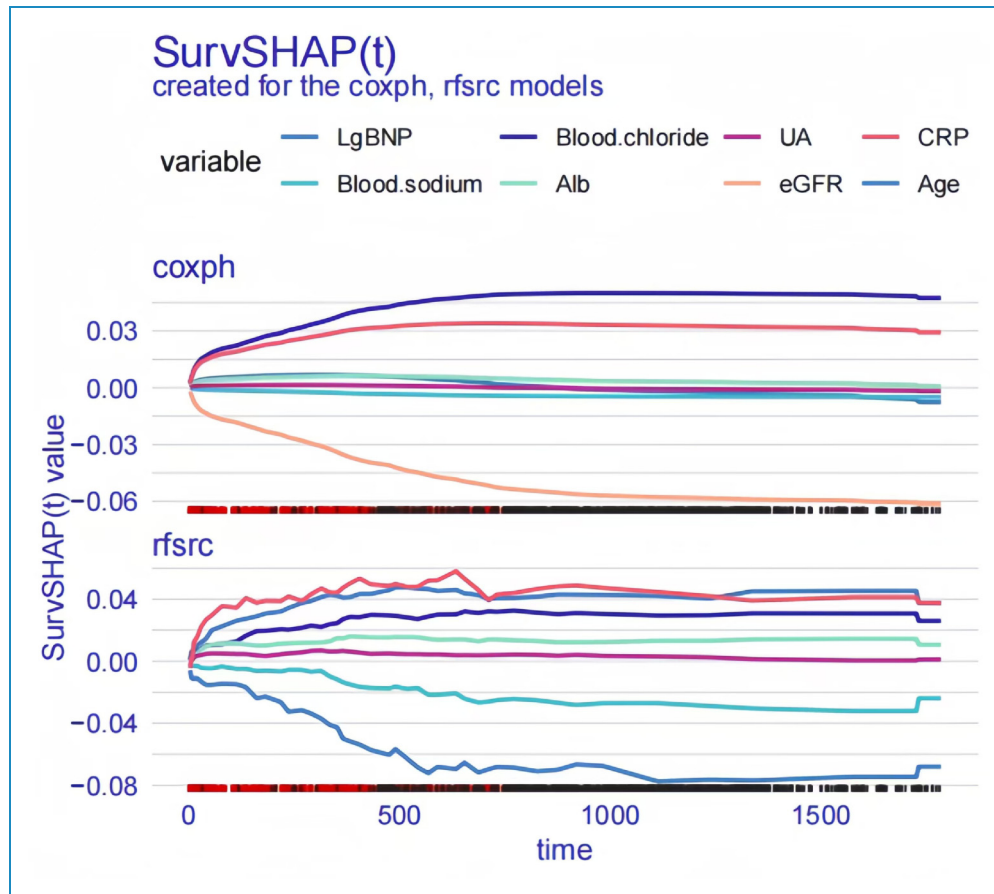


Figure 7. Local explanation: SurvSHAP(t) plot for a single patient. Alb: albumin; BNP: brain natriuretic peptide; coxph: Cox proportional hazards regression; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; rfsrc: random survival forest; UA: uric acid.

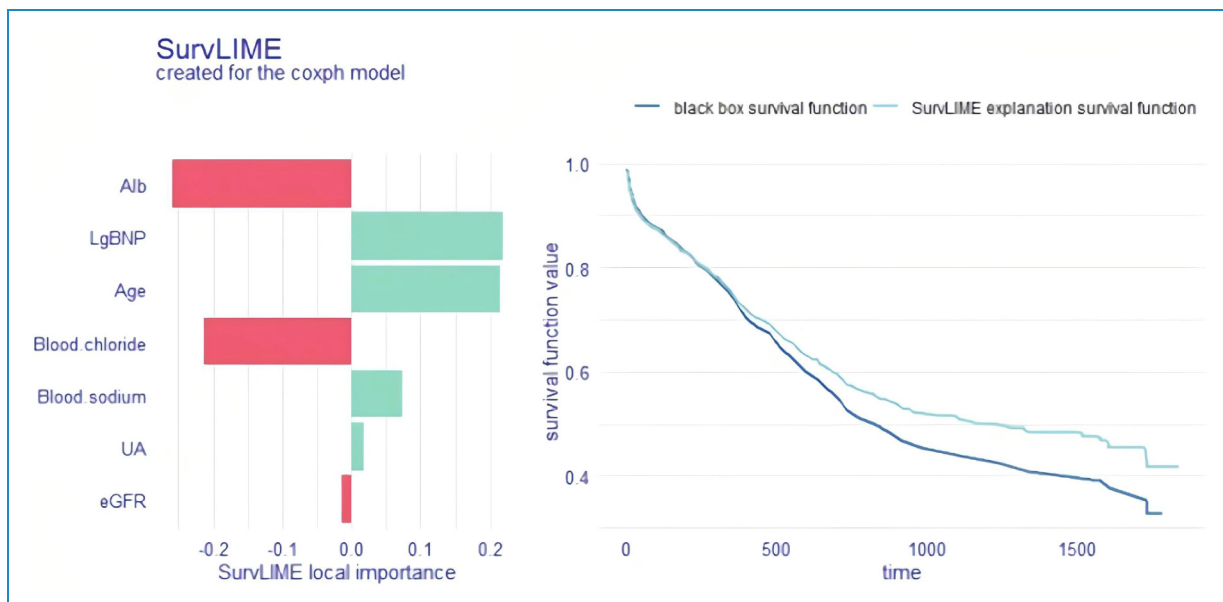


Figure 8. Local explanation: SurvLIME plot for a single patient; coxph model. Alb: albumin; BNP: brain natriuretic peptide; coxph: Cox proportional hazards regression; eGFR: estimated glomerular filtration rate; UA, uric acid.

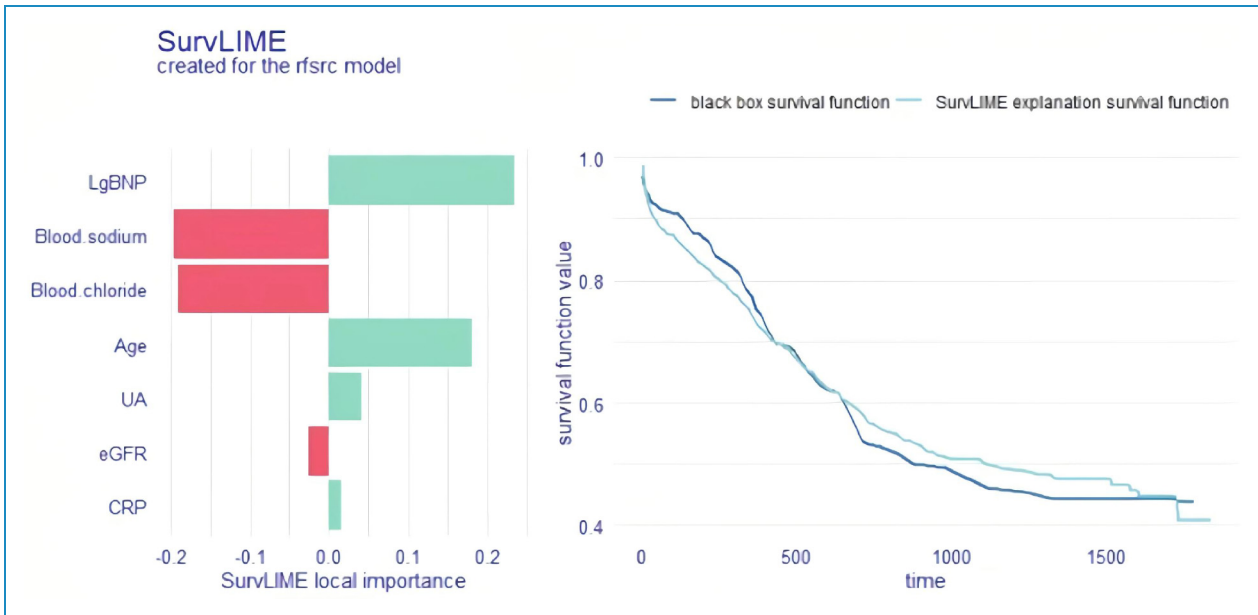


Figure 9. Local explanation: SurvLIME plot for a single patient; rfsrc model. BNP: brain natriuretic peptide; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; rfsrc: random survival forest; UA: uric acid.

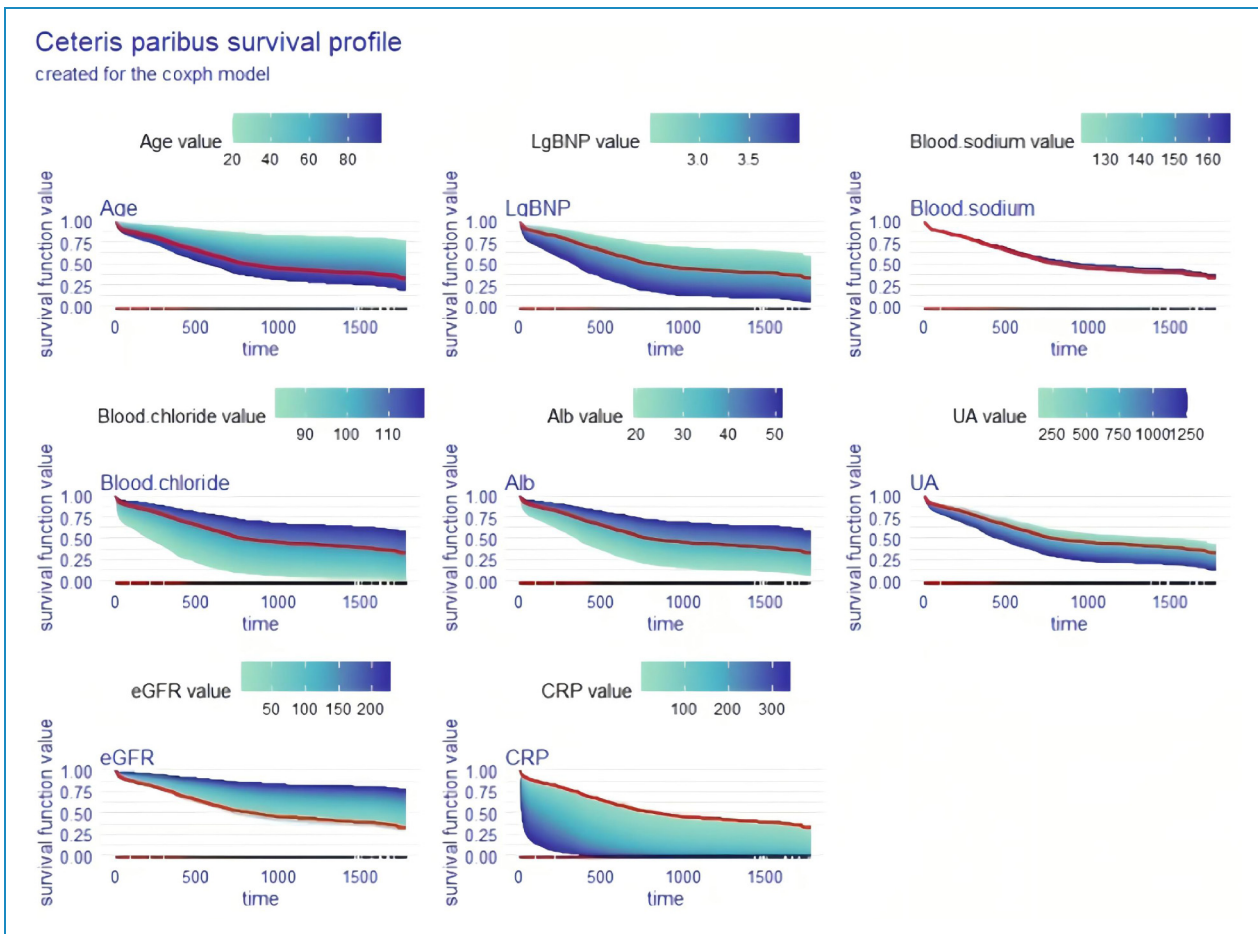


Figure 10. Local explanation: CPP for a single patient; coxph model. Alb: albumin; BNP: brain natriuretic peptide; coxph: Cox proportional hazards regression; CRP: C-reactive protein; CPP: ceteris paribus survival profile; eGFR: estimated glomerular filtration rate; UA: uric acid.

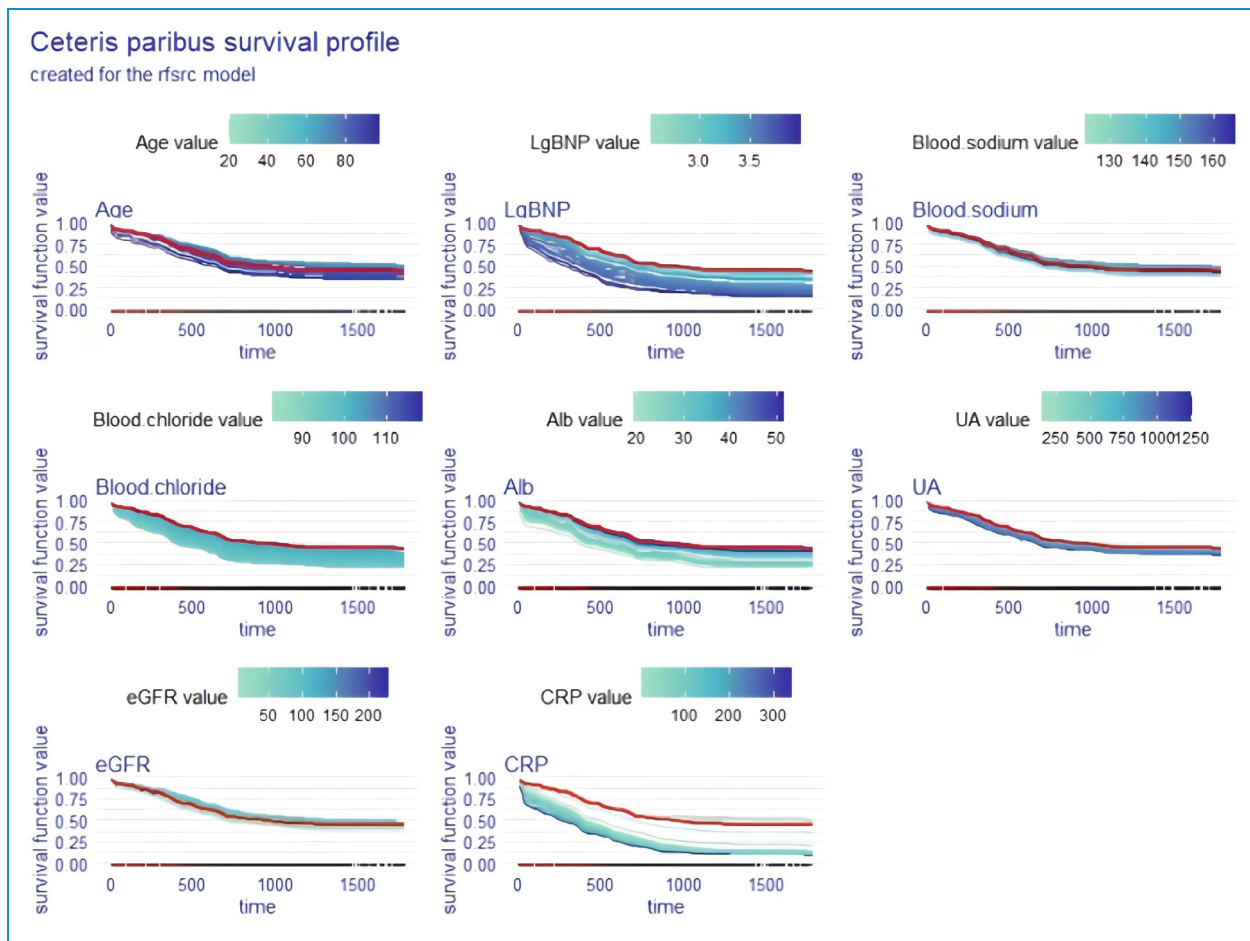


Figure 11. Local explanation: CPP for a single patient; rfsrc model. Alb: albumin; BNP: brain natriuretic peptide; CPP: ceteris paribus survival profile; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; rfsrc: random survival forest; UA: uric acid.

could be performed to validate the predictive models. Second, the subjects we chose were HF patients in NYHA class III or IV; thus, the prediction models might not be applicable to patients with milder HF symptoms, and we may gather information from patients in NYHA class I or II to improve the use of these prediction models in the future. Finally, this study ignored the issues of model validation and overfitting, and our later research will validate the model by dividing the population into a training set and a validation set to strengthen the model's credibility and prevent model overfitting.

Conclusion

The survex package in R not only allows for a more complete and intuitive interpretation of survival models but also allows for comparison and interpretation of results from different survival modelling methods. This study showed that the predictive model constructed by the rfsrc algorithm comprehensively outperformed the predictive model constructed by the coxph algorithm. Our models

can be used for both the entire population and for individual selected patients, and they can help clinicians personalise treatment plans based on the factors that affect the OS of each HF patient.

Contributorship: TS and LC researched literature and conceived the study. JY, NZ, WR, LG, PX, JZ, NZ and FY were involved in protocol development, gaining ethical approval, patient recruitment and data analysis. TS wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: The study was carried out in accordance with the Declaration of Helsinki and approved by the Medical Ethics Committee of the First Affiliated Hospital of Kunming Medical University. The ethics approval of the study was (2022) Ethics L No. 173.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Applied Basic Research Program of the Science and Technology Hall of Yunnan Province and Kunming Medical University (grant number 202301AY070001-130).

Guarantor: LC

ORCID iD: Lixing Chen  <https://orcid.org/0000-0002-0746-6706>

References

- Lippi G and Sanchis-Gomar F. Global epidemiology and future trends of heart failure. *AME Med J* 2020; 5: 15–15.
- Fu R, Xiang J, Bao H, et al. Association between process indicators and in-hospital mortality among patients with chronic heart failure in China. *Eur J Public Health* 2015; 25: 373–378.
- Moreno-Sánchez PA. Improvement of a prediction model for heart failure survival through explainable artificial intelligence. *Front Cardiovasc Med* 2023; 10: 1219586.
- Apley DW and Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. *J R Stat Soc Ser B: Stat Methodol* 2020; 82: 1059–1086.
- Molnar C. Iml: an R package for interpretable machine learning. *J Open Source Software* 2018; 3: 786.
- Di Martino F and Delmastro F. Explainable AI for clinical and remote health applications: a survey on tabular and time series data. *Artif Intell Rev* 2023; 56: 5261–5315.
- Baniecki H, Parzych D and Biecek P. The grammar of interactive explanatory model analysis. *Data Min Knowl Discov* 2023; 14: 1–37.
- Anon. Explainable Machine Learning in Survival Analysis. Available at: <https://modeloriented.github.io/survex/>. Accessed December 29, 2023.
- Spytek M, Krzyżiński M, Langbein SH. survex : an R package for explaining machine learning survival models. *Bioinformatics* 2023; 39: btad723.
- Kovalev MS and Utkin LV. A robust algorithm for explaining unreliable machine learning survival models using the Kolmogorov–Smirnov bounds. *Neural Netw* 2020; 132: 1–18.
- Cox DR. Regression models and life-tables. *J R Stat Soc Series B Methodol* 1972; 34: 187–202.
- Taylor JMG. Random survival forests. *J Thorac Oncol* 2011; 6: 1974–1975.
- Gotsman I, Shauer A, Zwas DR, et al. Low serum albumin: a significant predictor of reduced survival in patients with chronic heart failure. *Clin Cardiol* 2019; 42: 365–372.
- Yamada T, Haruki S, Minami Y, et al. The C-reactive protein to prealbumin ratio on admission and its relationship with outcome in patients hospitalized for acute heart failure. *J Cardiol* 2021; 78: 308–313.
- Yuan X, Huang B, Wang R, et al. The prognostic value of advanced lung cancer inflammation index (ALI) in elderly patients with heart failure. *Front Cardiovasc Med* 2022; 9: 934551.
- Sciomer S, Moscucci F, Salvioni E, et al. Role of gender, age and BMI in prognosis of heart failure. *Eur J Prev Cardiol* 2020; 27: 46–51.
- Burger PM, Koudstaal S, Mosterd A, et al. C-reactive protein and risk of incident heart failure in patients with cardiovascular disease. *J Am Coll Cardiol* 2023; 82: 414–426.
- Omar HR and Guglin M. Extremely elevated BNP in acute heart failure: patient characteristics and outcomes. *Int J Cardiol* 2016; 218: 120–125.
- Balling L, Schou M, Videbaek L, et al. Prevalence and prognostic significance of hyponatraemia in outpatients with chronic heart failure. *Eur J Heart Fail* 2011; 13: 968–973.
- Zhang Y, Peng R, Li X, et al. Serum chloride as a novel marker for adding prognostic information of mortality in chronic heart failure. *Clin Chim Acta* 2018; 483: 112–118.
- Tedeschi A, Agostoni P, Pezzuto B, et al. Role of comorbidities in heart failure prognosis part 2: chronic kidney disease, elevated serum uric acid. *Eur J Prev Cardiol* 2020; 27: 35–45.
- Mathur P, Srivastava S, Xu X, et al. Artificial intelligence, machine learning, and cardiovascular disease. *Clin Med Insights: Cardiol* 2020; 14: 1179546820927404.
- Ayers B, Sandholm T, Gosev I, et al. Using machine learning to improve survival prediction after heart transplantation. *J Card Surg* 2021; 36: 4113–4120.
- Xu C, Li H, Yang J, et al. Interpretable prediction of 3-year all-cause mortality in patients with chronic heart failure based on machine learning. *BMC Med Inform Decis Mak* 2023; 23: 267.
- Shapley LS. A value for n-person games. In: *17. A value for n-person games*. Princeton University Press, 2016, pp.307–318. Available at: <https://www.degruyter.com/document/doi/10.1515/9781400881970-018/html> (accessed June 2, 2024).
- Krzyżiński M, Spytek M, Baniecki H, et al. SurvSHAP(t): time-dependent explanations of machine learning survival models. *Knowl Based Syst* 2023; 262: 110234.
- Pachón-García C, Hernández-Pérez C, Delicado P, et al. SurvLIMEpy: a Python package implementing SurvLIME. *Expert Syst Appl* 2024; 237: 121620.