**Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP**

**Supplementary Material**

**1.Optimal** **Hyper-parameter**

The optimal hyper-parameters of each model were determined by Grid Search with 5-fold CV (Supplementary Table 1).

Supplementary Table 1 The optimal hyper-parameters of ML models determined by Grid Search with 5-fold CV.

|  |  |  |
| --- | --- | --- |
| Model | Parameter optimization range | Parameter |
| LR | 'C':[0.001,0.01,0.1,1,10,100],'penalty':['l1','l2','elasticnet','none'],'solver':['newton-cg','lbfgs', 'liblinear', 'sag', 'saga'] | C=0.1,penalty='l1', solver='liblinear', |
| KNN | n\_neighbors:(1, 15) | n\_neighbors=2 |
| SVM | 'C':(0.001,0.002,0.005,0.01,0.015,0.02,0.1,1,10,11,15,20),'gamma':(0.0001,0.001,0.002,0.005,0.01,0.015,0.02,0.1,1,10) | C=1,gamma=10 |
| NB | 'alpha':(0.5,0.6,0.7,0.8,0.9,1.0,1.1,1.2,1.3,1.4,1.5) | alpha=1.0 |
| MLP | 'hidden\_layer\_sizes':(5,10,15,20,25,30,35,40,45,50,),'learning\_rate\_init':np.arange(0.001,0.1,0.001) | hidden\_layer\_sizes=(10,), learning\_rate\_init=0.1 |
| XGBoost | 'n\_estimators':range(50,2000,10),'learning\_rate':np.arange(0.01,1,0.01),'max\_depth':range(1,10,1),'min\_child\_weight':range(1,10,1), 'gamma':[i/10.0for i in range(0, 5)], 'subsample':[i/10.0for i in range(3, 10)],'colsample\_bytree':[i/10.0 for i in range(6,10)], 'scale\_pos\_weight':[i for i in range(1,10,1)],'reg\_alpha':[1e-5,1e-2,0.1,1,2,2.5,3],'reg\_lambda':[1e-5,1e-2,0.1,1,2,2.5,3] | learning\_rate=0.03,n\_estimators=1600,max\_depth=7,min\_child\_weight=1,  gamma=0.3,subsample=0.6,reg\_alpha=3,reg\_lambda=2,colsample\_bytree=0.4,  objective='binary:logistic',nthread=4, scale\_pos\_weight=9 |  |

**2.Evaluation indexes**

F1-score is the weighted harmonic average of precision and recall, and is an indicator of the validity of the classification results by combining precision and recall. AUC is the area under the ROC curve. The ROC curve is a curve with (1−specificity) as the horizontal axis and sensitivity as the vertical axis. The calculation of AUC requires comprehensive consideration of sensitivity and specificity. The Brier score is a “calibrated” measure of a set of probabilistic predictions and is the mean squared difference between the observed and predicted results. Brier scores range from 0 to 1.00, with 0 representing the best possible calibration.

**3.****Feature Selection**

**3.1Candidate Variables**

The patient information was collected according to the case report form of chronic heart failure (CHF-CRF) developed by our research group according to the case record content and HF guidelines.CHF-CRF included the patient’s demographic(e.g., sex, smoking, drinking), vital signs (e.g., height, weight, temperature, blood oxygen saturation), heart failure causes (e.g., coronary heart disease, myocardial infarction, angina pectoris, obesity), heart failure complications (e.g., diabetes, arrhythmia, renal insufficiency, sleep apnea syndrome), symptoms and signs (e.g., palpitations, chest suppress, difficulty breathing), laboratory examination (e.g., WBC, RBC, hemoglobin, NT-proBNP), auxiliary examination, (e.g., electrocardiogram, dynamic electrocardiogram), and other examinations within 24 hours after admission and treatments (electrocardiogram monitoring and blood pressure monitoring), drug therapy (e.g., diuretics, angiotensin converting enzyme inhibitors, angiotensin receptor blockers, beta blockers), and oxygen therapy.

**3.2Single factor screening**

Continuous variables are reported as median (inter-quartile range) and compared with the wilcoxon rank sum test. Categorical variables are reported as frequencies (in percent) and compared with the chi-square test, P-value0.05 was considered statistically significant. The initially selected variables were shown in Supplementary Table 2.

Supplementary Table 2 Single-factor screening results and names of reserved variables and their assignments

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Full name** | **Unit /Assignment** | **Z/** | **P-value** |
| Event | --- | 0:Survival  1: No survival |  | --- |
| AGE | Age | years old | -11.846 | <0.001 |
| TEMP | Body temperature | degrees Celsius | -2.106 | 0.035 |
| DBP | Diastolic pressure | mmHg | -3.660 | <0.001 |
| WEIGHT | Weight | kg | -3.805 | <0.001 |
| BMI | Body mass index | kg/m2 | -4.499 | <0.001 |
| HR | Heart rate | times / minute | -5.376 | <0.001 |
| RBC | Red blood cells | 10^12/L | -5.409 | <0.001 |
| RDW | Red blood cell volume distribution width | % | -7.972 | <0.001 |
| HGB | Hemoglobin | g/L | -5.559 | <0.001 |
| NEU | Neutrophil absolute value | 10^10/L | -3.376 | 0.001 |
| N. | Neutrophil ratio | % | -7.051 | <0.001 |
| ALT | Alanine aminotransferase | U/L | -2.818 | 0.005 |
| ALB | Albumin | g/L | -3.806 | <0.001 |
| TBIL | Total bilirubin | μmol/L | -2.943 | 0.003 |
| DBIL | Direct bilirubin | μmol/L | -5.078 | <0.001 |
| X.GT | r-glutamyl transpeptidase | U/L | -1.941 | 0.052 |
| TG | Triglyceride | mmol/L | -4.590 | <0.001 |
| HDLC | High density lipoprotein cholesterol | μmol/L | -1.957 | 0.050 |
| LP(a) | Lipoprotein (a) | mg/L | -2.096 | 0.036 |
| BUN | Blood urea nitrogen | mmol/L | -8.629 | <0.001 |
| CR | Creatinine | mmol/L | -8.724 | <0.001 |
| UA | Uric acid | μmol/L | -6.632 | <0.001 |
| K | Potassium ions concentration | mmol/L | -3.467 | 0.001 |
| NA | Sodium ions concentration | mmol/L | -2.703 | 0.007 |
| CL | Chloride ion concentration | mmol/L | -3.996 | <0.001 |
| CYSC | Cystatin C | mg/L | -9.446 | <0.001 |
| NTPROBNP | N-terminal pronatriuretic peptide | ng/L | -9.898 | <0.001 |
| QRS | part of electrocardiographic wave time | ms | -3.558 | <0.001 |
| QTC | QT interval time | ms | -5.941 | <0.001 |
| LA | Left atrial front and rear diameter | mm | -7.857 | <0.001 |
| RA | Anteroposterior diameter of right atrium | mm | -5.576 | <0.001 |
| RA1 | Left and right diameter of right atrium | mm | -5.143 | <0.001 |
| LVDD | Internal diameter of left ventricular end-diastolic | mm | -4.287 | <0.001 |
| EF | Left ventricular ejection fraction | % | -7.438 | <0.001 |
| OCCUPATION | Occupation | 0:Manual worker,  1:Mental worker | 19.022 | <0.001 |
| HEALTHCARE | Types of health insurance | 0:Urban medical insurance,1:new rural cooperative medical care,2:full public expense,3:self-paying,4:other | 28.853 | <0.001 |
| MARITALSTA | Marital status | 1:unmarried,2:married,3:divorced,4: widowed | 11.533 | 0.009 |
| OMI | Old myocardial infarction | 0:no,1:yes | 5.494 | 0.019 |
| PCI | Percutaneous transluminal coronary intervention | 0:no,1:yes | 11.526 | 0.001 |
| CABG | Coronary artery bypass graft | 0:no,1:yes | 14.047 | <0.001 |
| NYHA  classification | New York Heart Association clinical classifications | 0:Ⅱ,  1:Ⅲ~Ⅳ | 57.424 | <0.001 |
| DRINKING | Drinking history | 0:never,1:ever | 5.671 | 0.017 |
| ALLERGY | Allergy history | 0:no,1:yes | 3.734 | 0.053 |
| ARRHYTHMIA | Atrial fibrillation | 0:no,1:yes | 40.912 | <0.001 |
| HLP | Hyperlipidemia | 0:no,1:yes | 15.586 | <0.001 |
| HISTORYOF0 | Central nervous system disease treatment history | 0:no,1:yes | 8.673 | 0.003 |
| PUMONARY | Pulmonary disease | 0:no,1:yes | 61.234 | <0.001 |
| RENAL | Renal insufficiency | 0:no,1:yes | 51.456 | <0.001 |
| CANSER | Cancer | 0:no,1:yes | 6.190 | 0.013 |
| DYSCRASIA | Dystrophy and sarcopenia | 0:no,1:yes | 4.605 | 0.032 |
| SAS | Sleep disorders and sleep breathing disorders | 0:no,1:yes | 28.785 | <0.001 |
| K.1 | Hyperkalemia/hypokalemia | 0:no,1:yes | 3.666 | 0.056 |
| IDA | Iron deficiency and anemia | 0:no,1:yes | 13.882 | <0.001 |
| SHORTNESS | Shortness of breath | 0:no,1:yes | 15.198 | <0.001 |
| ORTHOPNEA | Orthopnea | 0:no,1:yes | 35.297 | <0.001 |
| PAROXYSMAL | Paroxysmal nocturnal dyspnea | 0:no,1:yes | 61.514 | <0.001 |
| COUGH | Cough | 0:no,1:yes | 23.936 | <0.001 |
| LOSSOFAPPE | Loss of appetite | 0:no,1:yes | 46.141 | <0.001 |
| LACKOFENER | Lack of energy | 0:no,1:yes | 49.576 | <0.001 |
| ANKLEEDEMA | Ankle edema | 0:no,1:yes | 50.152 | <0.001 |
| PERIPHERAL | Peripheral edema | 0:no,1:yes | 45.111 | <0.001 |
| HEARTMURMU | Heart murmur | 0:no,1:yes | 14.251 | <0.001 |
| LR | Lung rales | 1:no,2: moist rales,  3:dry rales | 82.473 | <0.001 |
| NUTRITION | Nutrition | 1:bad,2: medium,  3: good,4: over | 11.602 | 0.003 |
| POSTURE | Posture | 1:autonomy,2:semi-recumbent position,  3: forced sitting | 6.969 | 0.008 |
| APPETITE | Appetite | 0: normal,1: decrease | 23.739 | <0.001 |
| E/A | Max left ventricle filling volume/Max left atrium filling volume | 0:<1  1:1~2  2:>2 | 10.947 | 0.004 |
| PE | Pericardial effusion | 1:no,2:little,  3: moderate,4: massive | 9.285 | 0.002 |
| WMA | Wall motion abnormalities | 0:no,1:yes | 18.855 | <0.001 |
| MR | Mitral regurgitation | 1:no,2:little,  3:moderate,4: massive | 68.156 | <0.001 |
| TR | Tricuspid regurgitation | 1:no,2: little,  3: moderate,4: massive | 51.373 | <0.001 |
| PVS1AI | Pulmonary aortic valve regurgitation | 1:no,2: little  3: moderate,4: massive | 50.133 | <0.001 |
| PVI | Pulmonary valve regurgitation | 1:no,2: little,  3: moderate,4: massive | 7.408 | 0.025 |
| AS | Aortic stenosis | 1:no,2: mild,  3: mild,4: severe | 8.089 | 0.004 |
| XRAY | X ray | 0:no,1:yes | 10.132 | 0.006 |
| AF | Atrial fibrillation | 0:no,1:paroxysmal,2:persistent | 30.228 | <0.001 |
| ST | Abnormal ST-T in electrocardiogram | 0:no,1:yes | 6.119 | 0.013 |
| INFECTION | lung infection | 0:no,1:yes | 71.573 | <0.001 |
| EDEMA | Pulmonary edema | 0:no,1:yes | 8.215 | 0.016 |
| Antifreezing | Use anticoagulant drugs | 0:no,1:yes | 15.507 | <0.001 |
| Statins | Use statins | 0:no,1:yes | 18.476 | <0.001 |
| Nitrates | Use nitrates | 0:no,1:yes | 5.158 | 0.023 |