Supplementary Materials

Establishment and validation of a machine learning model to stratify malnutrition risk in hospitalized older patients with chronic heart failure

Qiuhong Sun BD[1] and Jing Che BD[2]

[1]*Department of Cardiology, The First Hospital of China Medical University, Liaoning, China*
[2]*Blood Collection Center, The First Hospital of China Medical University, Liaoning, China*

Corresponding Author:
Dr Jing Che, Blood Collection Center, The First Hospital of China Medical University, No. 155, Nanjing North Street, 110001, Heping District, Shenyang City, Liaoning Province, China.
Tel: +86-02483308172.
Email: chejing@cmu1h.com
Manuscript received 27 November 2024. Initial review completed 28 June 2025. Revision accepted 09 September 2025.
doi: 10.6133/apjcn.202602_35(1).0014

Supplementary Table 1. Brief introduction of the machine learning algorithms

Algorithms	Description	Strengths	Weakness
LR	A linear model used for binary classification that estimates the probability of a categorical dependent variable.	Simple, interpretable, effective for linearly separable data	Not suitable for complex relationships
NNet	Neural networks are computational models inspired by human brain structure, used for tasks like classification, regression, and more. They consist of layers of interconnected nodes (neurons) that process input data.	Can model complex patterns, supports large datasets, highly versatile for different tasks	Requires large amounts of data, computationally expensive, can be hard to interpret
SVM	A classification method that finds the hyperplane best separating different classes in a high-dimensional space.	Effective in high-dimensional spaces, robust to overfitting	Requires careful parameter tuning, poor with large datasets
NB	A probabilistic classifier based on Bayes' theorem, assuming independence among predictors.	Fast, works well with small datasets and text classification	Assumes independence, less accurate with large datasets
RF	An ensemble method using multiple decision trees to improve accuracy and control overfitting through majority voting or averaging.	High accuracy, robust to overfitting, handles missing values	Slower to train, less interpretable
XGB	An advanced implementation of gradient boosting designed for speed and performance, often used in structured/tabular data competitions.	High predictive power, flexibility, efficient computation	Complex, requires careful tuning
LGBM	A gradient boosting framework that grows tree leaf-wise and uses histogram-based optimization for faster computation.	Fast training speed, high efficiency with large datasets	Can be sensitive to overfitting
CAT	A gradient boosting algorithm that handles categorical features directly and reduces overfitting by applying ordered boosting.	Excels with categorical features, less prone to overfitting	Complexity, requires parameter tuning

LR, Logistic Regression; SVM, Support Vector Machine; NNet, Neural Network; NB, Naïve Bayes; RF, Random Forest; XGB; Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; CAT, CATBoost

Supplementary Table 2. Hyperparameter tuning process for the machine learning model

Algorithms	Hyperparameter tuning	R package
LR	Alpha: Regularization type weight (L1/L2 mix ratio): 0.1, 0.5, 1.0 (search range) Lambda: Regularization strength: 0.01, 0.05, 0.1, 0.2 (search range); Thresh: Convergence threshold: 1e-5 (fixed value); Maxit: Maximum number of iterations: 1000 (fixed value).	glm
RF	Mtry: Number of features randomly selected per tree: 2, 3 (search range); Ntree: Number of decision trees: 200, 300 (search range); Nodesize: Minimum samples in leaf node: 45, 55 (search range); Maxnodes: Maximum number of nodes per tree: 22, 26 (search range); Sampsize: Sample size per tree: 380 (fixed value); Replace: Whether to sample with replacement: TRUE (fixed value); Strata: Whether to stratify sampling (balance class distribution): TRUE (fixed value); Importance: Whether to compute feature importance: TRUE (fixed value); Proximity: Whether to compute sample proximity: TRUE (fixed value).	randomForest
XGB	Max_depth: Maximum depth of trees: 1, 2 (search range); Eta: Learning rate: 0.01 (fixed value); Subsample: Sample proportion for training each tree: 0.4, 0.5 (search range); Colsample_bytree: Feature sample proportion for training each tree: 0.4, 0.5 (search range); Reg_alpha: L1 regularization strength: 0.3 (fixed value); Reg_lambda: L2 regularization strength: 3.0, 3.5 (search range); Gamma: Minimum loss reduction required for node splitting: 0.3, 0.5 (search range); Nrounds: Number of iterations (number of trees): 300 (fixed value); Objective: Type of loss function: “binary:logistic” (fixed value); Stratified: Whether cross-validation is stratified: TRUE (fixed value); Early_stopping_rounds: Early stopping rounds (prevent overfitting): 30 (fixed value); Weight: Sample weights (balance classes): positive = 1.1, negative = 1 (fixed value).	xgboost
NNet	Size: Number of neurons in hidden layer: 3, 5 (search range); Decay: Weight decay (regularization strength): 0.1, 0.2, 0.3 (search range); Maxit: Maximum number of iterations: 500 (fixed value); Rang: Weight initialization range: 0.05 (fixed value); MaxNWts: Maximum number of weights (prevent model complexity): 10000 (fixed value); Skip: Whether to skip direct connection from input to output layer: FALSE (fixed value); Softmax: Whether to use softmax activation (disabled for binary classification): FALSE (fixed value); Trace: Whether to output training log: FALSE (fixed value).	nnet

LR, Logistic Regression; SVM, Support Vector Machine; NNet, Neural Network; NB, Naïve Bayes; RF, Random Forest; XGB; Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; CAT, CATBoost

Supplementary Table 2. Hyperparameter tuning process for the machine learning model (cont.)

Algorithms	Hyperparameter tuning	R package
LGBM	Num_leaves: Number of leaf nodes: 10, 14 (search range); Learning_rate: Learning rate: 0.008, 0.012 (search range); Reg_lambda: L2 regularization strength: 3.0, 4.0 (search range); Subsample: Sample proportion for training each tree: 0.6 (fixed value); Colsample_bytree: Feature sample proportion for training each tree: 0.6 (fixed value); Objective: Type of loss function: “binary” (fixed value); Metric: Evaluation metric: “auc” (fixed value); Min_data_in_leaf: Minimum data in leaf node: 30 (fixed value); Min_sum_hessian_in_leaf: Minimum sum hessian in leaf (control leaf purity): 25 (fixed value); Reg_alpha: L1 regularization strength: 0.2 (fixed value); Verbose: Log output level: -1 (fixed value, no output); Early_stopping_rounds: Early stopping rounds: 50 (fixed value).	LightGBM
SVM	C: Penalty coefficient (regularization strength): 0.1, 1, 10 (search range); Sigma: Kernel bandwidth (radial basis kernel): 0.01, 0.1, 1 (search range); Method: Kernel type: “svmRadial” (fixed value, radial basis kernel); PreProcess: Data preprocessing method: c(“center”, “scale”) (fixed value); Probability: Whether to output prediction probability: TRUE (fixed value); Metric: Evaluation metric: “ROC” (fixed value).	e1071
NB	Laplace: Laplace smoothing coefficient: 0, 1, 2 (search range); Usekernel: Whether to use kernel density estimation (alternative to parameter estimation): FALSE, TRUE (search range); Adjust: Adjustment parameter for kernel density estimation: 1.2, 1.5, 2 (search range); PreProcess: Data preprocessing method: c(“center”, “scale”) (fixed value); Metric: Evaluation metric: “ROC” (fixed value).	naiveBayes
CAT	Depth: Maximum depth of trees: 1, 3 (search range); Iterations: Number of iterations (number of trees): 200, 300 (search range); Learning_rate: Learning rate: 0.02, 0.05 (search range); Loss_function: Type of loss function: “Logloss” (fixed value); Eval_metric: Evaluation metric: “AUC” (fixed value); Verbose: Log output level: 0 (fixed value, no output); Prediction_type: Prediction type (Probability / Category): “Probability” (fixed value).	CatBoost

LR, Logistic Regression; SVM, Support Vector Machine; NNet, Neural Network; NB, Naïve Bayes; RF, Random Forest; XGB; Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; CAT, CATBoost

Supplementary Table 3. Clinical characteristics between patients at training and test sets (n = 790)

Parameters	Total (n = 790)	Training set (n = 553)	Test set (n = 237)	Statistics	*p* value
Age	74.0 [67.0, 80.0]	73.0 [67.0, 80.0]	74.0 [67.0, 81.0]	-0.752	0.452
Malnutrition					
No	442 (55.9%)	302 (54.6%)	140 (59.1%)	1.34	0.247
Yes	348 (44.1%)	251 (45.4%)	97 (40.9%)		
Gender					
Men	346 (43.8%)	234 (42.3%)	112 (47.3%)	1.65	0.199
Women	444 (56.2%)	319 (57.7%)	125 (52.7%)		
Residence					
Rural (%)	349 (44.2%)	244 (44.1%)	105 (44.3%)	0.002	0.963
Urban (%)	441 (55.8%)	309 (55.9%)	132 (55.7%)		
Education level					
< high school	518 (65.6%)	362 (65.5%)	156 (65.8%)	0.010	0.922
≥high school	272 (34.4%)	191 (34.5%)	81 (34.2%)		
Monthly household income (Yuan)					
<3000	353 (44.7%)	259 (46.8%)	94 (39.7%)	3.86	0.145
3000-5000	293 (37.1%)	200 (36.2%)	93 (39.2%)		
>5000	144 (18.2%)	94 (17.0%)	50 (21.1%)		
BMI (kg/m^2)	24.2 [20.6, 27.4]	24.2 [20.5, 27.5]	24.2 [20.6, 27.4]	-0.071	0.943
Current smoking (%)	298 (37.7%)	217 (39.2%)	81 (34.2%)	1.81	0.178
Current drinking (%)	294 (37.2%)	212 (38.3%)	82 (34.6%)	0.992	0.319
Disease duration (months)					
≤6	594 (75.2%)	420 (75.9%)	174 (73.4%)	0.570	0.450
>6	196 (24.8%)	133 (24.1%)	63 (26.6%)		
NYHA classification					
II	113 (14.3%)	77 (13.9%)	36 (15.2%)	0.769	0.681
III	403 (51.0%)	279 (50.5%)	124 (52.3%)		
IV	274 (34.7%)	197 (35.6%)	77 (32.5%)		
Teeth number					
≥20	321 (40.6%)	224 (40.5%)	97 (40.9%)	0.012	0.912
<20	469 (59.4%)	329 (59.5%)	140 (59.1%)		
Living alone (%)	71 (9.0%)	48 (8.7%)	23 (9.7%)	0.213	0.644
Comorbidities					
Peripheral edema (%)	459 (58.1%)	321 (58.0%)	138 (58.2%)	0.002	0.962
Hypertension (%)	511 (64.7%)	360 (65.1%)	151 (63.7%)	0.140	0.709

PG- BMI, Body Mass Index; NYHA, New York Heart Association; COPD, Chronic Obstructive Pulmonary Disease; CHD, Coronary Heart Disease; CKD, Chronic Kidney Disease; MRA, Mineralocorticoid Receptor Antagonist; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitor/Angiotensin II Receptor Blocker; TP, Total Protein; HGB, Hemoglobin; Cr, Creatinine; Hs-CRP, High-sensitivity C-Reactive Protein; Na, Sodium; K, Potassium; FBG, Fasting Blood Glucose; NT-proBNP, N-Terminal pro B-type Natriuretic Peptide; WBC, White Blood Cell; Alb, Albumin; Neu, Neutrophil; Lym, Lymphocyte; PLT, Platelet; Mon, Monocyte; NLR, Neutrophil-to-Lymphocyte Ratio; PLR, Platelet-to-Lymphocyte Ratio; PNI, Prognostic Nutritional Index.
Continuous values are presented as median [IQR] and category values are presented as frequency (%).

Supplementary Table 3. Clinical characteristics between patients at training and test sets (n = 790)

Parameters	Total (n = 790)	Training set (n = 553)	Test set (n = 237)	Statistics	*p* value
Comorbidities					
Dyslipidemia (%)	366 (46.3%)	259 (46.8%)	107 (45.1%)	0.190	0.663
COPD (%)	156 (19.7%)	105 (19.0%)	51 (21.5%)	0.671	0.413
CHD (%)	479 (60.6%)	332 (60.0%)	147 (62.0%)	0.275	0.600
Atrial fibrillation (%)	272 (34.4%)	185 (33.5%)	87 (36.7%)	0.779	0.378
Valvular heart disease (%)	211 (26.7%)	146 (26.4%)	65 (27.4%)	0.089	0.765
Diabetes (%)	227 (28.7%)	163 (29.5%)	64 (27.0%)	0.495	0.482
Anemia (%)	358 (45.3%)	253 (45.8%)	105 (44.3%)	0.140	0.708
CKD (%)	281 (35.6%)	198 (35.8%)	83 (35.0%)	0.044	0.833
Gastrointestinal disease (%)	301 (38.1%)	203 (36.7%)	98 (41.4%)	1.52	0.218
Functional capacity					
Hand grasp (kg)	17.0 [13.0, 21.0]	17.0 [13.0, 21.0]	17.0 [14.0, 21.0]	-0.693	0.488
Upper arm circumference (cm)	24.0 [20.0, 30.0]	25.0 [21.0, 30.0]	24.0 [19.0, 29.0]	-1.54	0.124
Calf circumference (cm)	34.0 [31.0, 37.0]	35.0 [31.0, 37.0]	34.0 [31.0, 37.0]	-1.08	0.280
Medication					
Diuretics (%)	536 (67.8%)	379 (68.5%)	157 (66.2%)	0.399	0.528
MRA (%)	678 (85.8%)	483 (87.3%)	195 (82.3%)	3.50	0.062
β-blocker (%)	438 (55.4%)	313 (56.6%)	125 (52.7%)	0.999	0.317
ACEI/ARB (%)	587 (74.3%)	406 (73.4%)	181 (76.4%)	0.758	0.384
Medical number (≥5)	306 (38.7%)	218 (39.4%)	88 (37.1%)	0.367	0.545
Laboratory tests					
TP (g/L)	54.0 [46.0, 63.0]	55.0 [46.0, 64.0]	54.0 [46.0, 62.5]	-0.973	0.331
HGB (g/L)	133 [121, 147]	133 [122, 149]	133 [121, 147]	-0.553	0.580
Cr (mg/dL)	1.50 [1.10, 1.90]	1.50 [1.10, 1.90]	1.50 [1.10, 1.90]	-0.180	0.857
Hs-CRP (<0.5mg/L)	503 (63.7%)	356 (64.4%)	147 (62.0%)	0.396	0.529
Na (mmol/L)	135 [129, 141]	135 [130 ,140]	135 [129,141]	-0.482	0.630
K (mmol/L)	4.10 [3.80, 4.50]	4.10 [3.80, 4.50]	4.10 [3.80, 4.50]	-0.612	0.541
FBG (g/L)	5.30 [4.70, 5.90]	5.20 [4.70, 6.00]	5.30 [4.70, 5.90]	-0.154	0.878
eGFR (ml/min/1.73m^2)	70.0 [56.0, 87.0]	71.0 [58.0, 87.0]	70.0 [55.0, 87.0]	-1.160	0.246
BNP (mg/L)	372 [222, 491]	373 [220, 491]	369 [224, 494]	-0.276	0.782
NT-proBNP (ng/L)	2989 [2032, 3938]	2998 [2050, 3916]	2986 [2020, 3948]	-0.089	0.929
Alb (mg/L)	36.0 [28.0, 41.0]	35.0 [27.0, 41.0]	36.0 [29.0, 41.0]	-1.73	0.084
WBC (10^9/L)	10.1 [8.40, 12.5]	10.2 [8.70, 12.6]	10.0 [8.30, 12.3]	-1.37	0.170

PG- BMI, Body Mass Index; NYHA, New York Heart Association; COPD, Chronic Obstructive Pulmonary Disease; CHD, Coronary Heart Disease; CKD, Chronic Kidney Disease; MRA, Mineralocorticoid Receptor Antagonist; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitor/Angiotensin II Receptor Blocker; TP, Total Protein; HGB, Hemoglobin; Cr, Creatinine; Hs-CRP, High-sensitivity C-Reactive Protein; Na, Sodium; K, Potassium; FBG, Fasting Blood Glucose; NT-proBNP, N-Terminal pro B-type Natriuretic Peptide; WBC, White Blood Cell; Alb, Albumin; Neu, Neutrophil; Lym, Lymphocyte; PLT, Platelet; Mon, Monocyte; NLR, Neutrophil-to-Lymphocyte Ratio; PLR, Platelet-to-Lymphocyte Ratio; PNI, Prognostic Nutritional Index.
Continuous values are presented as median [IQR] and category values are presented as frequency (%).

Supplementary Table 3. Clinical characteristics between patients at training and test sets (n = 790)

Parameters	Total (n = 790)	Training set (n = 553)	Test set (n = 237)	Statistics	*p* value
Neu (10^9/L)	8.00 [6.00, 11.0]	8.00 [5.00, 10.0]	8.00 [6.00, 11.0]	-1.28	0.200
Lym (10^9/L)	1.70 [1.40, 2.00]	1.70 [1.40, 2.00]	1.70 [1.30, 2.00]	-0.286	0.775
PLT (10^9/L)	190 [145, 246]	187 [144, 243]	192 [147, 248]	-1.02	0.307
Mon (10^9/L)	0.52 [0.28, 0.73]	0.49 [0.26, 0.72]	0.56 [0.31, 0.73]	-1.54	0.123
NLR	4.71 [3.16, 6.67]	4.76 [3.33, 6.67]	4.62 [2.95, 6.43]	-1.01	0.314
PLR	111 [81.8, 155]	108 [82.0, 147]	112 [81.5, 156]	-0.709	0.478
PNI	44.5 [38.0, 50.0]	44.0 [37.5, 49.3]	45.0 [38.5, 50.0]	-1.54	0.123

PG- BMI, Body Mass Index; NYHA, New York Heart Association; COPD, Chronic Obstructive Pulmonary Disease; CHD, Coronary Heart Disease; CKD, Chronic Kidney Disease; MRA, Mineralocorticoid Receptor Antagonist; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitor/Angiotensin II Receptor Blocker; TP, Total Protein; HGB, Hemoglobin; Cr, Creatinine; Hs-CRP, High-sensitivity C-Reactive Protein; Na, Sodium; K, Potassium; FBG, Fasting Blood Glucose; NT-proBNP, N-Terminal pro B-type Natriuretic Peptide; WBC, White Blood Cell; Alb, Albumin; Neu, Neutrophil; Lym, Lymphocyte; PLT, Platelet; Mon, Monocyte; NLR, Neutrophil-to-Lymphocyte Ratio; PLR, Platelet-to-Lymphocyte Ratio; PNI, Prognostic Nutritional Index.
Continuous values are presented as median [IQR] and category values are presented as frequency (%).

Supplementary Table 4. Clinical characteristics between patients at development and external validation sets (n = 1128)

Parameters	Total (n = 1128)	External validation set (n = 338)	Development set (n = 790)	Statistics	*p* value
Age	74.0 [67.0, 80.0]	74.0 [67.8, 81.0]	74.0 [67.0, 80.0]	-0.543	0.587
Malnutrition					
No	643 (57.0%)	201 (59.5%)	442 (55.9%)	1.20	0.274
Yes	485 (43.0%)	137 (40.5%)	348 (44.1%)		
Gender					
Men	496 (44.0%)	150 (44.4%)	346 (43.8%)	0.032	0.857
Women	632 (56.0%)	188 (55.6%)	444 (56.2%)		
Residence					
Rural (%)	487 (43.2%)	138 (40.8%)	349 (44.2%)	1.08	0.298
Urban (%)	641 (56.8%)	200 (59.2%)	441 (55.8%)		
Education level					
< high school	754 (66.8%)	236 (69.8%)	518 (65.6%)	1.93	0.165
≥high school	374 (33.2%)	102 (30.2%)	272 (34.4%)		
Monthly household income (Yuan)					
<3000	503 (44.6%)	150 (44.4%)	353 (44.7%)	0.834	0.659
3000-5000	412 (36.5%)	119 (35.2%)	293 (37.1%)		
>5000	213 (18.9%)	69 (20.4%)	144 (18.2%)		
BMI (kg/m^2)	24.1 [20.3, 27.4]	23.4 [20.0, 27.2]	24.2 [20.6, 27.4]	-1.06	0.290
Current smoking (%)	417 (37.0%)	119 (35.2%)	298 (37.7%)	0.642	0.423
Current drinking (%)	415 (36.8%)	121 (35.8%)	294 (37.2%)	0.204	0.651
Disease duration (months)					
≤6	844 (74.8%)	250 (74.0%)	594 (75.2%)	0.189	0.664
>6	284 (25.2%)	88 (26.0%)	196 (24.8%)		
NYHA classification					
II	161 (14.3%)	48 (14.2%)	113 (14.3%)	0.006	0.997
III	575 (51.0%)	172 (50.9%)	403 (51.0%)		
IV	392 (34.8%)	118 (34.9%)	274 (34.7%)		
Teeth number					
≥20	457 (40.5%)	136 (40.2%)	321 (40.6%)	0.015	0.901
<20	671 (59.5%)	202 (59.8%)	469 (59.4%)		
Living alone (%)	101 (9.0%)	30 (8.9%)	71 (9.0%)	0.004	0.952
Comorbidities					
Peripheral edema (%)	657 (58.2%)	198 (58.6%)	459 (58.1%)	0.022	0.881
Hypertension (%)	734 (65.1%)	223 (66.0%)	511 (64.7%)	0.174	0.677

PG- BMI, Body Mass Index; NYHA, New York Heart Association; COPD, Chronic Obstructive Pulmonary Disease; CHD, Coronary Heart Disease; CKD, Chronic Kidney Disease; MRA, Mineralocorticoid Receptor Antagonist; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitor/Angiotensin II Receptor Blocker; TP, Total Protein; HGB, Hemoglobin; Cr, Creatinine; Hs-CRP, High-sensitivity C-Reactive Protein; Na, Sodium; K, Potassium; FBG, Fasting Blood Glucose; NT-proBNP, N-Terminal pro B-type Natriuretic Peptide; WBC, White Blood Cell; Alb, Albumin; Neu, Neutrophil; Lym, Lymphocyte; PLT, Platelet; Mon, Monocyte; NLR, Neutrophil-to-Lymphocyte Ratio; PLR, Platelet-to-Lymphocyte Ratio; PNI, Prognostic Nutritional Index.
Continuous values are presented as median [IQR] and category values are presented as frequency (%). *$p < 0.05$.

Supplementary Table 4. Clinical characteristics between patients at development and external validation sets (n = 1128)

Parameters	Total (n = 1128)	External validation set (n = 338)	Development set (n = 790)	Statistics	*p* value
Comorbidities					
Dyslipidemia (%)	521 (46.2%)	155 (45.9%)	366 (46.3%)	0.021	0.884
COPD (%)	216 (19.1%)	60 (17.8%)	156 (19.7%)	0.609	0.435
CHD (%)	688 (61.0%)	209 (61.8%)	479 (60.6%)	0.144	0.705
Atrial fibrillation (%)	373 (33.1%)	101 (29.9%)	272 (34.4%)	2.21	0.137
Valvular heart disease (%)	316 (28.0%)	105 (31.1%)	211 (26.7%)	2.23	0.136
Diabetes (%)	324 (28.7%)	97 (28.7%)	227 (28.7%)	0.001	0.990
Anemia (%)	506 (44.9%)	148 (43.8%)	358 (45.3%)	0.224	0.636
CKD (%)	391 (34.7%)	110 (32.5%)	281 (35.6%)	0.957	0.328
Gastrointestinal disease (%)	404 (35.8%)	103 (30.5%)	301 (38.1%)	5.99	0.014*
Functional capacity					
Hand grasp (kg)	17.0 [13.0, 21.0]	16.0 [13.0, 21.0]	17.0 [13.0, 21.0]	-1.33	0.184
Upper arm circumference (cm)	25.0 [20.0, 29.0]	25.0 [20.0, 29.0]	24.0 [20.0, 30.0]	-0.220	0.826
Calf circumference (cm)	34.0 [31.0, 37.0]	34.0 [31.0, 37.0]	34.0 [31.0, 37.0]	-0.456	0.649
Medication					
Diuretics (%)	776 (68.8%)	240 (71.0%)	536 (67.8%)	1.10	0.294
MRA (%)	968 (85.8%)	290 (85.8%)	678 (85.8%)	0.001	0.992
β-blocker (%)	606 (53.7%)	168 (49.7%)	438 (55.4%)	3.14	0.077
ACEI/ARB (%)	838 (74.3%)	251 (74.3%)	587 (74.3%)	0.002	0.988
Medical number (≥5)	443 (39.3%)	137 (40.5%)	306 (38.7%)	0.321	0.571
Laboratory tests					
TP (g/L)	54.0 [46.0, 63.0]	55.0 [46.0, 62.0]	54.0 [46.0, 63.0]	-0.243	0.808
HGB (g/L)	133.0 [121.0, 148.0]	133.0 [121.0, 148.0]	133.0 [121.0, 147.0]	-0.339	0.734
Cr (mg/dL)	1.50 [1.10, 2.10]	1.60 [1.20, 2.00]	1.50 [1.10, 1.90]	-2.01	0.045*
Hs-CRP (<0.5mg/L)	723 (64.1%)	220 (65.1%)	503 (63.7%)	0.207	0.649
Na (mmol/L)	135 [129, 141]	134 [129, 140]	135 [129, 141]	-0.956	0.339
K (mmol/L)	4.10 [3.80, 4.50]	4.20 [3.80, 4.43]	4.10 [3.80, 4.50]	-0.569	0.569
FBG (g/L)	5.30 [4.60, 5.90]	5.20 [4.60, 5.90]	5.30 [4.70, 5.90]	-0.365	0.715
eGFR (ml/min/1.73m^2)	71.0 [56.0, 86.0]	73.0 [56.0, 84.3]	70.0 [56.0, 87.0]	-0.244	0.807
BNP (mg/L)	370 [220, 491]	353 [212, 493]	372 [222, 491]	-0.357	0.721
NT-proBNP (ng/L)	2981 [2049, 3938]	2974 [2118, 3942]	2989 [2032, 3938]	-0.350	0.726
Alb (mg/L)	36.0 [28.0, 41.0]	36.0 [28.0, 41.0]	36.0 [28.0, 41.0]	-0.669	0.503
WBC (10^9/L)	10.1 [8.3, 12.5]	10.0 [8.3, 12.6]	10.1 [8.4, 12.5]	-0.088	0.930

PG- BMI, Body Mass Index; NYHA, New York Heart Association; COPD, Chronic Obstructive Pulmonary Disease; CHD, Coronary Heart Disease; CKD, Chronic Kidney Disease; MRA, Mineralocorticoid Receptor Antagonist; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitor/Angiotensin II Receptor Blocker; TP, Total Protein; HGB, Hemoglobin; Cr, Creatinine; Hs-CRP, High-sensitivity C-Reactive Protein; Na, Sodium; K, Potassium; FBG, Fasting Blood Glucose; NT-proBNP, N-Terminal pro B-type Natriuretic Peptide; WBC, White Blood Cell; Alb, Albumin; Neu, Neutrophil; Lym, Lymphocyte; PLT, Platelet; Mon, Monocyte; NLR, Neutrophil-to-Lymphocyte Ratio; PLR, Platelet-to-Lymphocyte Ratio; PNI, Prognostic Nutritional Index.
Continuous values are presented as median [IQR] and category values are presented as frequency (%).
*$p < 0.05$.

Supplementary Table 4. Clinical characteristics between patients at development and external validation sets (n = 1128)

Parameters	Total (n = 1128)	External validation set (n = 338)	Development set (n = 790)	Statistics	*p* value
Neu (10^9/L)	8.00 [6.00, 11.0]	8.00 [6.00, 11.0]	8.00 [6.00, 11.0]	-0.610	0.542
Lym (10^9/L)	1.70 [1.40, 2.00]	1.70 [1.40, 2.00]	1.70 [1.40, 2.00]	-0.975	0.329
PLT (10^9/L)	189 [148, 247]	184 [152, 249]	190 [145, 246]	-0.230	0.818
Mon (10^9/L)	0.50 [0.27, 0.73]	0.47 [0.23, 0.76]	0.52 [0.28, 0.73]	-0.911	0.362
NLR	4.67 [3.16, 6.67]	4.46 [3.08, 6.67]	4.71 [3.16, 6.67]	-0.057	0.955
PLR	110 [80.9, 154]	109 [78.3, 153]	111 [81.8, 155]	-0.269	0.788
PNI	44.5 [38.0, 49.9]	44.0 [37.9, 49.6]	44.5 [38.0, 50.0]	-0.382	0.702

PG- BMI, Body Mass Index; NYHA, New York Heart Association; COPD, Chronic Obstructive Pulmonary Disease; CHD, Coronary Heart Disease; CKD, Chronic Kidney Disease; MRA, Mineralocorticoid Receptor Antagonist; ACEI/ARB, Angiotensin-Converting Enzyme Inhibitor/Angiotensin II Receptor Blocker; TP, Total Protein; HGB, Hemoglobin; Cr, Creatinine; Hs-CRP, High-sensitivity C-Reactive Protein; Na, Sodium; K, Potassium; FBG, Fasting Blood Glucose; NT-proBNP, N-Terminal pro B-type Natriuretic Peptide; WBC, White Blood Cell; Alb, Albumin; Neu, Neutrophil; Lym, Lymphocyte; PLT, Platelet; Mon, Monocyte; NLR, Neutrophil-to-Lymphocyte Ratio; PLR, Platelet-to-Lymphocyte Ratio; PNI, Prognostic Nutritional Index.
Continuous values are presented as median [IQR] and category values are presented as frequency (%).
*$p < 0.05$.

Supplementary Table 5. Multivariable logistic regression analysis

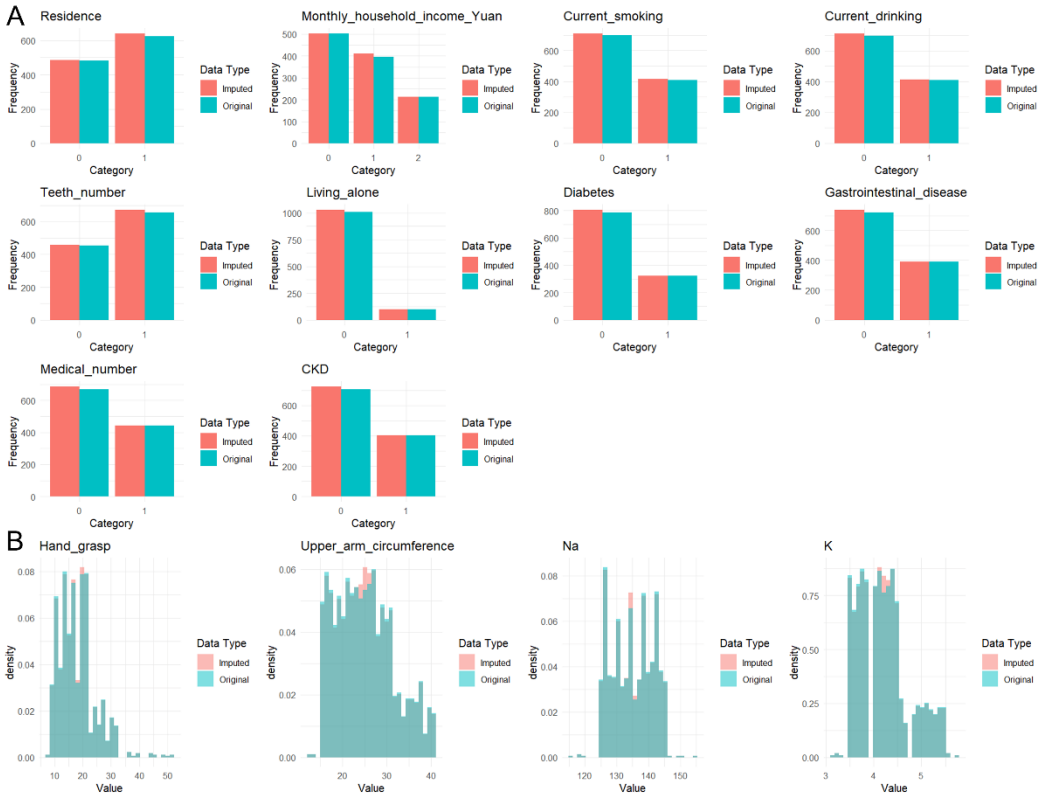
Variables	B	SE	OR	95% CI	Z	Exact *p*-value	p_{FDR}†
NYHA classification	1.11	0.173	3.03	2.15-4.25	6.39	<0.001***	<0.001***
Diabetes	1.42	0.238	4.12	2.58-6.57	5.95	<0.001***	<0.001***
Medical number	0.692	0.218	2.00	1.31-3.06	3.18	0.001**	0.002**
Hs-CRP	0.644	0.222	1.90	1.23-2.94	2.90	0.004**	0.004**
Age	0.058	0.013	1.06	1.03-1.09	4.56	<0.001***	<0.001***
BMI	-0.114	0.023	0.892	0.853-0.933	-5.03	<0.001***	<0.001***
PNI	-0.032	0.014	0.968	0.942-0.995	-2.32	0.020*	0.020*
Calf circumference	-0.173	0.031	0.841	0.791-0.894	-5.60	<0.001***	<0.001***

BMI, Body Mass Index; NYHA classification, New York Heart Association classification; Hs-CRP, High-sensitivity C-Reactive Protein; PNI, Prognostic Nutritional Index; B, regression coefficient; SE, standard error; OR, odds ratio; CI, confidence interval; Z, Z-score; FDR, false discovery rate.
†p-value adjusted for false discovery rate based on the eight variables in the current analysis
*p < 0.05, **p < 0.01, ***p < 0.001.

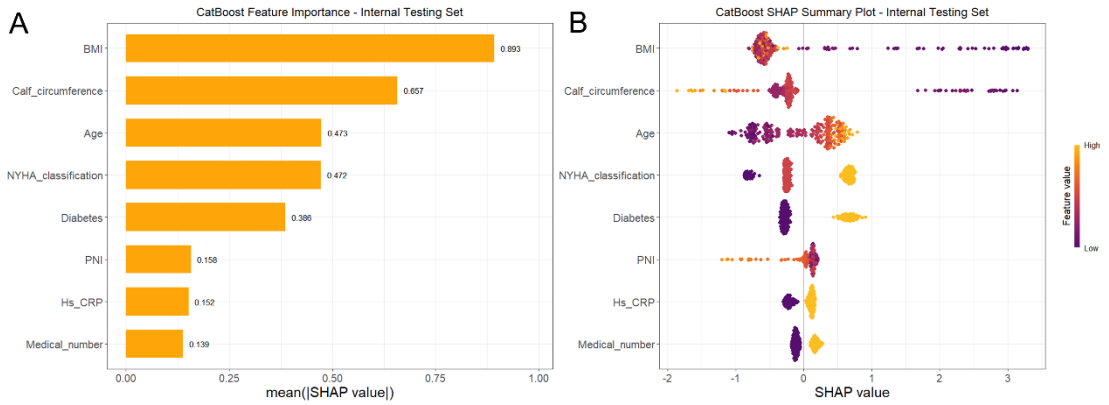
Supplementary Table 6. Performance of machine learning-based models in the training and test sets

Algorithms	Accuracy	AUC (95% CI)	Recall	Precision	F1 score	Specificity	NPV	Brier score
Training set								
LR	0.770	0.846 (0.814-0.878)	0.709	0.767	0.737	0.821	0.773	0.159
RF	0.835	0.914 (0.891-0.937)	0.721	0.896	0.799	0.931	0.801	0.120
XGB	0.821	0.911 (0.888-0.935)	0.673	0.909	0.774	0.944	0.777	0.140
LGBM	0.788	0.873 (0.844-0.901)	0.705	0.805	0.752	0.858	0.778	0.155
NNet	0.797	0.890 (0.865-0.916)	0.745	0.796	0.770	0.841	0.799	0.135
CAT	0.859	0.947 (0.930-0.964)	0.753	0.922	0.829	0.947	0.822	0.099
SVM	0.814	0.899 (0.874-0.925)	0.777	0.806	0.791	0.844	0.820	0.128
NB	0.799	0.901 (0.876-0.926)	0.665	0.861	0.751	0.911	0.766	0.130
Internal testing set								
LR	0.776	0.837 (0.783-0.892)	0.680	0.750	0.714	0.843	0.792	0.155
RF	0.827	0.887 (0.842-0.932)	0.732	0.826	0.776	0.893	0.828	0.127
XGB	0.819	0.883 (0.837-0.929)	0.680	0.846	0.754	0.914	0.805	0.143
LGBM	0.751	0.831 (0.779-0.884)	0.691	0.698	0.694	0.793	0.787	0.168
NNet	0.789	0.858 (0.807-0.909)	0.691	0.770	0.728	0.857	0.800	0.143
CAT	0.840	0.901 (0.858-0.943)	0.753	0.839	0.794	0.900	0.840	0.113
SVM	0.772	0.842 (0.789-0.896)	0.691	0.736	0.713	0.829	0.795	0.155
NB	0.810	0.898 (0.857-0.939)	0.680	0.825	0.746	0.900	0.803	0.125
External validation set								
LR	0.754	0.855 (0.816-0.893)	0.737	0.682	0.709	0.766	0.811	0.159
RF	0.822	0.900 (0.867-0.933)	0.759	0.794	0.776	0.866	0.841	0.124
XGB	0.846	0.908 (0.877-0.939)	0.737	0.863	0.795	0.920	0.837	0.140
LGBM	0.787	0.857 (0.817-0.897)	0.745	0.734	0.739	0.816	0.824	0.162
NNet	0.793	0.878 (0.842-0.914)	0.818	0.713	0.762	0.776	0.862	0.145
CAT	0.828	0.916 (0.887-0.945)	0.774	0.797	0.785	0.866	0.849	0.116
SVM	0.760	0.856 (0.817-0.895)	0.745	0.689	0.716	0.771	0.816	0.156
NB	0.802	0.891 (0.858-0.923)	0.693	0.792	0.739	0.876	0.807	0.133

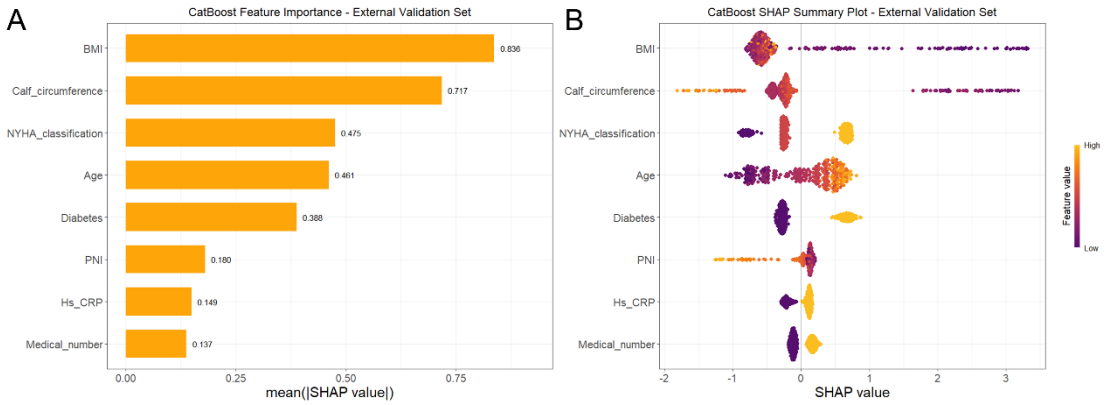
AUC, Area Under curve; CI, Confidence Interval; NPV, negative predictive value; LR, Logistic Regression; SVM, Support Vector Machine; NNet, Neural Network; NB, Naïve Bayes; RF, Random Forest; XGB; Extreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; CAT, CATBoost



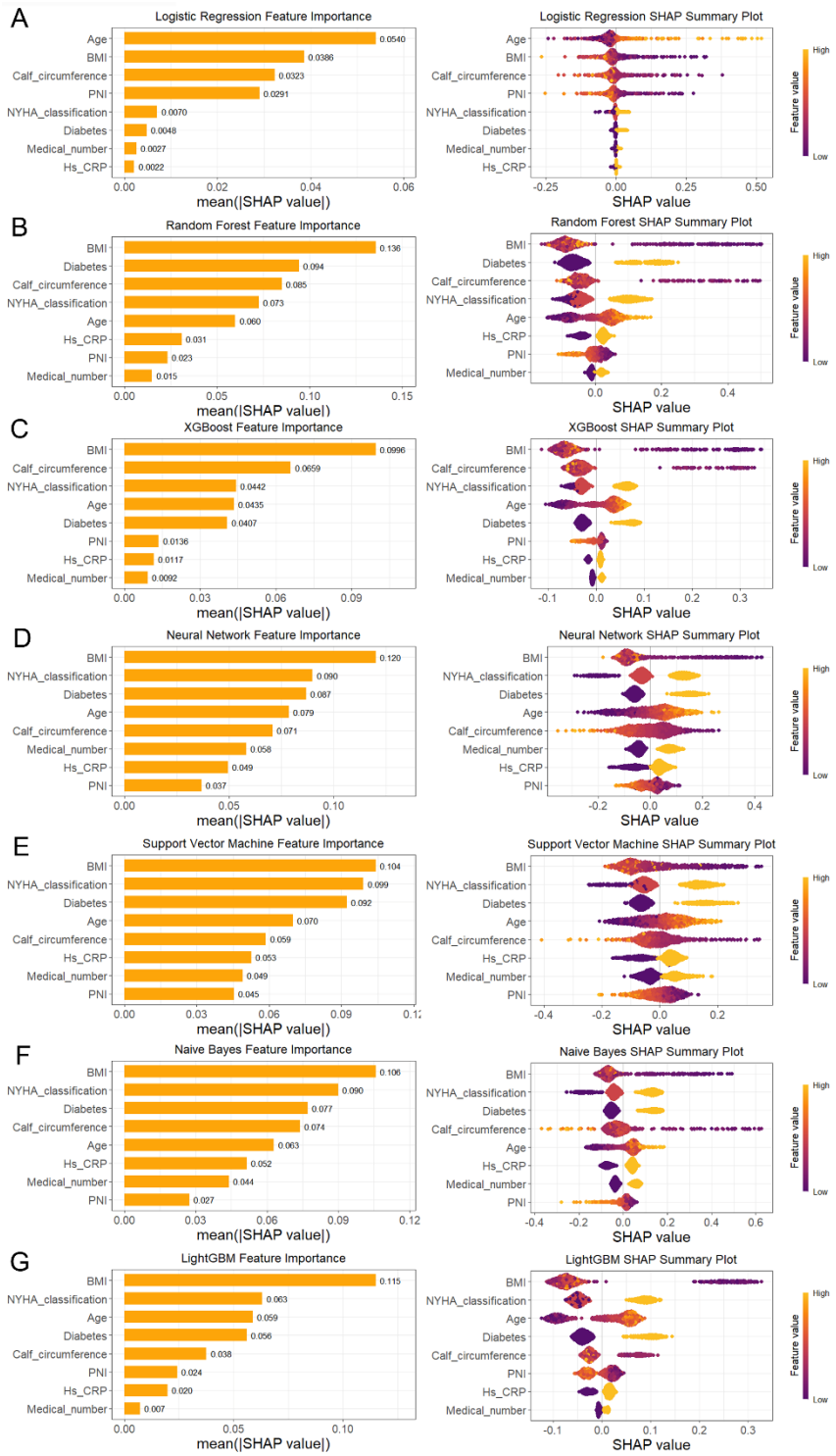
Supplementary Figure 1. Assessment of missing data patterns before and after imputation. (A) Distribution of missing values for category variables. (B) Distribution of missing values for continuous variables.



Supplementary Figure 2. SHAP analysis based on the CatBoost model in the internal testing set. (A) Summary plot of feature importance. (B) Beeswarm plot illustrating feature importance.



Supplementary Figure 3. SHAP analysis based on the CatBoost model in the external validation set. (A) Summary plot of feature importance. (B) Beeswarm plot illustrating feature importance.



Supplementary Figure 4. SHAP summary plot for the eight influential variables in the other predictive models. (A) Logistic regression. (B) Random Forest. (C) XGBoost. (D) Neural Network. (E) Support Vector Machine. (F) Naïve Bayes; (G) LightGBM.