Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) in press
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2500804

Machine learning models for predicting thirty-day mortality following TAVR: a national study from the NTCVR cohort

Qifeng ZHU[1,2,3,4*], Jin LU[1,2,3*], Jiayuan LI[1,2,3], Feiyu WU[1,2,3], Danqing YU[1,5], Yihan PAN[1], Qijing ZHOU[6], Chongzhou ZHENG[7], Daxin ZHOU[8,9], Wenzhi PAN[8,9], Xianbao LIU[1,2,3,4✉], Jian'an WANG[1,2,3,4 ✉]

[1]*Department of Cardiology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310009, China*
[2]*State Key Laboratory of Transvascular Implantation Devices, Hangzhou 310009, China*
[3]*Heart Regeneration and Repair Key Laboratory of Zhejiang Province, Hangzhou 310009, China*
[4]*Transvascular Implantation Devices Research Institute, Hangzhou 310053, China*
[5]*Department of Cardiology, Cixi Integrated Traditional Chinese and Western Medicine Medical and Health Group, Cixi 315300, China*
[6]*Department of Radiology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310009, China*
[7]*Department of Structural Heart Disease, Cardiovascular Medicine Center, Affiliated Hospital of Guangdong Medical University, Zhanjiang 524001, China*
[8]*Department of Cardiology, Zhongshan Hospital, Fudan University, Shanghai 200032, China*
[9]*National Clinical Research Center for Interventional Medicine, Shanghai 200032, China*

Abstract: Background: Transcatheter aortic-valve replacement (TAVR) has emerged as the preferred treatment for patients with aortic-valve stenosis (AS). However, risk assessment tools tailored to this patient group remain insufficient. Aims: We aimed to develop a machine-learning-based model for predicting thirty-day mortality risk in TAVR patients, using a national cohort from China. Methods: This multicenter, registry-based study (NTCVR) included 10,799 patients undergoing TAVR at 147 Chinese tertiary hospitals (November 2011 - August 2024). Patients were split into three sets: training (60%), internal validation (20%), and external validation (20%, two separate provinces). Oversampling addressed class imbalance (30-day mortality rate: 2.9%). Extensive feature selection and model development employed 15 feature subsets and 15 ML learners, generating 1,125 candidate models. SHapley Additive exPlanations (SHAP) analysis was used to further assess the influence of selected predictors and machine-learning models. Results: The optimal model combined Double Input Symmetrical Relevance (DISR) feature selection with a Support Vector Machine (SVM) learner. The Area Under the ROC Curve (AUC) of the optimal machine learning model was significantly higher than that of the Society of Thoracic Surgeons (STS) score in both the internal validation cohort [0.74 (0.67, 0.81) vs. 0.60 (0.53, 0.67), $p < 0.05$] and the external validation cohort [0.69 (0.62, 0.76) vs. 0.57 (0.51, 0.63), $p < 0.05$]. SHapley Additive exPlanations (SHAP) analysis of the variables included in the optimal model highlighted the contributions to mortality prediction of baseline alanine aminotransferase, creatinine, NYHA class, need for circulatory support, and non-elective procedure status. Conclusions: A simple and effective machine-learning-based model was developed for predicting thirty-day mortality in TAVR patients, offering a valuable tool for risk stratification in China.

Key words: Machine Learning; Transcatheter aortic-valve replacement (TAVR); Mortality; Predictive model

✉ Jian'an WANG, email: wangjianan111@zju.edu.cn;
✉ Xianbao LIU, email:liuxb@zju.edu.cn;
[*] The two authors contributed equally to this work
Jian'an WANG, https://orcid.org/0000-0002-4583-3204;
Xianbao LIU,https://orcid.org/0000-0003-1556-9198;

Received Dec. 7, 2025; Revision accepted Mar. 12, 2026;
Crosschecked xxx. xx, 20xx; Published online xxx. xx, 20xx

1 Introduction

With increasing life expectancy and the onset of an aging society, the incidence of aortic stenosis (AS) continues to rise.(Iung and Vahanian, 2012; Osnabrugge et al., 2013; Xu et al., 2022) In recent years, transcatheter aortic-valve replacement (TAVR) has emerged as a primary treatment for these patients.(Otto et al., 2021; Vahanian et al., 2022) Since 2019, TAVR has surpassed surgical aortic-valve replacement in volume, becoming the predominant therapy for calcific AS. (Carroll et al., 2020; Structural Heart Disease Group of Chinese College of Cardiovascular Physician, 2024)

Despite significant reductions in procedural complications due to advances in TAVR techniques and iterative device improvements, short-term adverse events remain a key concern as TAVR indications expand to include younger and lower-risk patients. (Arnold et al., 2024) Accurate preoperative risk stratification is essential for prognostic evaluation of TAVR candidates. However, current risk-assessment tools, such as the widely used Society of Thoracic Surgeons (STS) score, are not specifically tailored for TAVR and tend to overestimate thirty-day mortality post-TAVR due to their Surgical Aortic Valve Replacement (SAVR) -based nature.(Beohar et al., 2014; Kumar et al., 2018) Furthermore, mainstream risk models, including STS and the European System for Cardiac Operative Risk Evaluation (EuroSCORE), were primarily developed using Western populations, which raises concerns about their direct applicability in predicting mortality in Chinese TAVR patients.(O'brien et al., 2009; Nashef et al., 2012; Jilaihawi et al., 2015; Chen et al., 2023) Thus, the development of population-specific predictive models for prognostic evaluation in Chinese cohorts is crucial.

Machine learning, a subset of artificial intelligence, offers distinct advantages in analyzing complex relationships between multiple features and disease outcomes, and is used for a growing number of applications in medical diagnosis and outcome prediction.(Deo, 2015; Cai et al., 2024) However, there is limited data on using machine-learning techniques to enhance the predictive power of statistical models for TAVR outcomes. Consequently, our goal in this study was to develop and validate machine-learning models for predicting thirty-day mortality following TAVR.

2 Methods

2.1 Study Design

This was a multicenter and retrospective cohort study that included data from 147 hospitals across China, spanning a period from November 2011 to August 2024. The study adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence (TRIPOD + AI) guidelines(Collins et al., 2024).

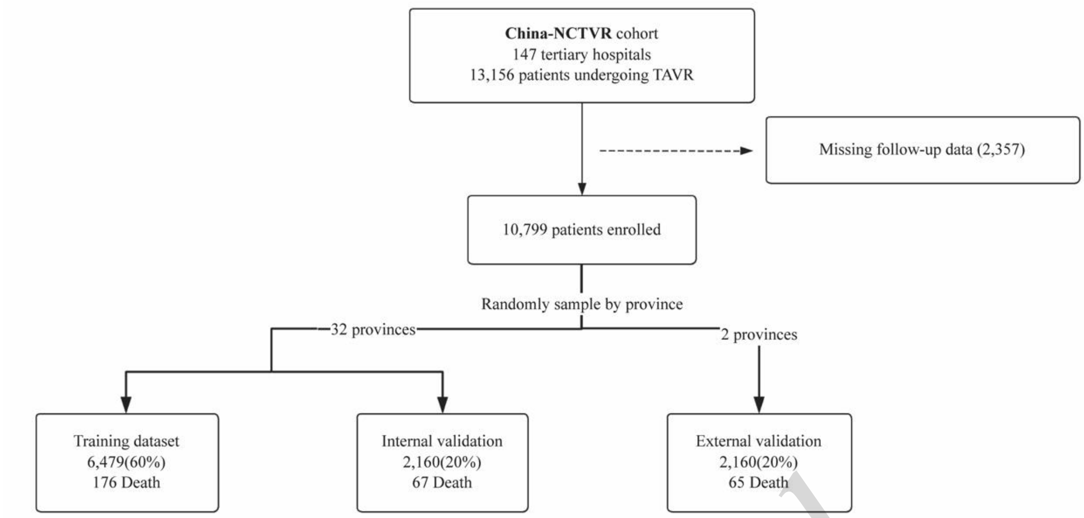

Fig. 1 Flow Diagram of Sample Selection and Split. In the China-NCTVR cohort, 13,156 patients who underwent TAVR from Feb 2017 to Nov 2022 were initially selected. After excluding 2,357 patients without complete outcome data during follow-up, 10,799 patients were included in the final analysis. Patients from two randomly selected provinces were assigned to the external validation set (2,160 patients, approximately 20%), while those from the remaining 32 provinces were divided into a training set (6,479 patients, approximately 60%) and internal validation set (2,160 patients, approximately 20%). NCTVR, The National Transcatheter Valve Therapeutics Registry.

2.2 Data Source and Patient Population

The National Transcatheter Valve Therapeutics Registry (NTCVR), the largest database of TAVR procedures performed in China, was co-initiated by the Chinese Cardiovascular Association (CCA) and the National Clinical Research Center for Interventional Medicine.(Hong et al., 2022) All patients in this cohort were aged 65 or older and primarily underwent TAVR due to severe AS or aortic regurgitation. A total of 10,799 patients were included in the development of the prediction model after excluding those lacking complete 1-month follow-up data (n = 2,357) (Fig. 1). Clinical data were obtained from the electronic medical-record system.

2.3 Outcomes and Predictors

Thirty-day mortality post-TAVR was defined as the primary outcome, with Approximately 40 preoperative predictors selected from the NTCVR database (Table S1). These variables included patient demographics, clinical features, comorbidities, lab-test results, and imaging characteristics (e.g., CT, echocardiogram). Intraoperative and postoperative variables were excluded to focus on predicting endpoint events based solely on preoperative data.

2.4 Data Preprocessing

The initial step in the data-preprocessing phase was excluding obvious outliers, which were considered data-entry errors; followed by refinement using the 3-sigma rule. Special attention was given to resolving issues arising from discrepancies in measurement units across hospitals and conducting internal consistency checks among interrelated clinical indicators. Of the 40 variables, the majority exhibited missing values in fewer than 5% of the total sample. However, certain variables such as coronary height, STS score, and calcification severity displayed higher missing rates, likely due to these parameters not being mandatorily assessed during preoperative evaluations at participating centers. We handled missing baseline data using multiple imputation with chained equations under the missing-at-random assumption, generating five imputed datasets with five iterations(Sterne et al., 2009). Imputation was performed on the full dataset before dataset partitioning. After imputation, the completed dataset was divided based on provinces. We randomly selected two provinces as the external validation cohort, and split the remaining data into training and internal validation sets. The final proportions for the training set, internal validation set, and external validation set were approximately 6:2:2. To address class imbalance in the endpoint events, we applied oversampling using the Synthetic Minority Oversampling Technique (SMOTE) only to the training set after dataset splitting. The internal and external validation sets retained their original class distributions and were not subjected to any resampling procedures.

2.5 Statistical Analysis

Five feature-selection methods were utilized to assess the importance of variables in predicting outcomes: Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Lasso, Anova t-score, and Double Input Symmetrical Relevance (DISR). The top 15 variables were selected based on their significance in the feature ranking. Fifteen machine-learning algorithms were then applied, including Generalized Linear Model, RF, XGBoost, Iterative Dichotomiser 3, Classification and Regression Trees, C4.5 Algorithm, C5.0 Algorithm, Conditional Inference Trees, Interactive Dichotomizer Algorithm, K-Nearest Neighbors, Backpropagation, Least Absolute Shrinkage and Selection Operator, Multi-Layer Perceptron, Gradient Boosting Machine, and Decision Tree. The models were trained on the training dataset, with hyperparameters optimized through cross-validation using grid search. We assessed model performance on the internal validation set using standard metrics such as accuracy, area under the curve, sensitivity, specificity, and accuracy. The final selected model was compared to the traditional STS score. To explain the model's predictions, we used SHapley Additive exPlanations (SHAP) values to evaluate the contribution of individual variables to the risk prediction (Fig. S1). All machine-learning algorithms, statistical analyses, and figures were implemented using R software, version 4.2.1. The full code is available upon reasonable request.

3 Results

3.1 Cohort characteristics

A total of 10,799 TAVR patients were included in the final analysis. The mean age was 75.03 years, with 6,429 males (59.5%) in the cohort. Thirty-day mortality occurred in 308 patients (2.9%). The median LVEF was 58.0%, and valve calcification was observed in 7,839 cases (78.2%). Circulatory support was planned and utilized in 309 patients (2.9%). Baseline characteristics are detailed in Table 1.

Table 1 Baseline characteristics of patients undergoing TAVR stratified by datasets

Variables	Overall (n=10799)	Training dataset (n=7906)	Internal validation (n=1977)	External validation (n=916)
Age (mean (SD))	75·03 (6·23)	74·83 (6·12)	75·20 (6·41)	76·48 (6·48)
Gender (%)	6429 (59·5)	4695 (59·4)	1218 (61·6)	516 (56·3)
BMI (median [IQR])	22·78 [20·58, 25·01]	22·83 [20·69, 25·00]	22·67 [20·52, 25·03]	22·72 [20·45, 25·08]
Symptoms (%)	9929 (92·0)	7303 (92·4)	1820 (92·2)	806 (88·0)
Dyslipidemia (%)	1774 (16·4)	1289 (16·3)	331 (16·8)	154 (16·8)
Hypertension (%)	5800 (53·7)	4192 (53·0)	1057 (53·5)	551 (60·2)
Diabetes (%)	1992 (18·4)	1436 (18·2)	396 (20·1)	160 (17·5)
AV block (%)	673 (6·2)	474 (6·0)	121 (6·1)	78 (8·5)
Hemoglobin (mean (SD))	122·81 (19·94)	123·03 (19·70)	122·39 (21·13)	121·84 (19·28)
Creatinine (median [IQR])	0·90 [0·75, 1·13]	0·90 [0·75, 1·12]	0·90 [0·75, 1·14]	0·90 [0·74, 1·12]
ALT (median [IQR])	18·00 [12·00, 27·00]	18·00 [12·00, 27·30]	17·00 [12·00, 27·00]	18·00 [13·00, 26·00]
NYHA class (%)				
I	468 (4·3)	358 (4·5)	78 (3·9)	32 (3·5)
II	3112 (28·8)	2344 (29·6)	590 (29·9)	178 (19·4)
III	5553 (51·4)	3960 (50·1)	1000 (50·6)	593 (64·7)
IV	1664 (15·4)	1244 (15·7)	307 (15·5)	113 (12·3)
STS (median [IQR])	4·73 [2·88, 7·00]	4·78 [2·95, 7·00]	5·00 [3·00, 7·35]	4·00 [2·36, 6·24]
Reasons for TAVR (%)				
Aortic Stenosis	9340 (86·5)	6878 (87·0)	1717 (86·9)	745 (81·4)
Pure Aortic Regurgitation	1340 (12·4)	945 (12·0)	238 (12·1)	157 (17·2)
Surgical aortic valve reintervention	96 (0·9)	70 (0·9)	16 (0·8)	10 (1·1)
TAVR valve reintervention	19 (0·2)	12 (0·2)	4 (0·2)	3 (0·3)
Annulus perimeter (mean (SD))	77·76 (8·56)	77·75 (8·52)	77·72 (8·85)	77·98 (8·26)
Annulus area (mean (SD))	469·61 (105·51)	469·67 (105·27)	470·34 (108·25)	467·45 (101·47)
Ascend aorta (mean (SD))	37·31 (5·52)	37·33 (5·46)	37·22 (5·62)	37·38 (5·86)
Left coronary height (mean (SD))	13·89 (3·67)	13·85 (3·71)	14·03 (3·67)	13·86 (3·24)
Right coronary height (mean (SD))	16·19 (3·73)	16·21 (3·74)	16·24 (3·76)	15·89 (3·57)
Aortic angulation (mean (SD))	52·06 (11·01)	52·02 (10·98)	51·72 (11·13)	53·75 (10·92)
Valve calcification (%)	7839 (78·2)	5776 (78·4)	1456 (79·0)	607 (74·5)
Valve calcification level (%)				
Degree0	1137 (18·6)	845 (18·7)	186 (16·9)	106 (22·0)
Degree1	1529 (25·0)	1151 (25·4)	290 (26·4)	88 (18·3)
Degree2	1372 (22·5)	1046 (23·1)	256 (23·3)	70 (14·5)
Degree3	1605 (26·3)	1188 (26·3)	289 (26·3)	128 (26·6)
Degree4	462 (7·6)	294 (6·5)	78 (7·1)	90 (18·7)
Annulus calcification (%)	2453 (40·2)	1806 (39·9)	435 (39·6)	212 (44·0)
LVOT calcification (%)	916 (15·0)	664 (14·7)	159 (14·5)	93 (19·3)
LVIDd (mean (SD))	52·36 (9·50)	52·46 (9·59)	52·24 (9·44)	51·71 (8·80)
LVPWd (median [IQR])	12·00 [10·00, 13·00]	12·00 [10·00, 13·00]	12·00 [10·00, 13·00]	11·00 [10·00, 13·00]
IVSd (mean (SD))	12·53 (2·31)	12·55 (2·31)	12·53 (2·24)	12·39 (2·41)

LVEF (median [IQR])	58·00 [47·40, 64·40]	58·00 [47·00, 64·00]	58·00 [47·00, 64·00]	60·00 [50·60, 65·00]
AR (%)	9055 (83·9)	6633 (83·9)	1640 (83·0)	782 (85·4)
AR level (%)				
None/Trace	899 (9·9)	610 (9·2)	143 (8·7)	146 (18·7)
Mild	3154 (34·8)	2348 (35·4)	571 (34·8)	235 (30·1)
Moderate	2424 (26·8)	1787 (26·9)	462 (28·2)	175 (22·4)
Severe	2577 (28·5)	1887 (28·5)	464 (28·3)	226 (28·9)
MR level (%)				
None/Trace	1241(14·1)	825(12·9)	229(14·1)	187(24·4)
Mild	4461(50·8)	3260(51·0)	849(52·3)	352(46·0)
Moderate	2134(24·3)	1598(25·0)	367(22·6)	169(22·1)
Moderate-to-Severe	465(5·3)	352(5·5)	85(5·2)	28(3·7)
Severe	484(5·5)	362(5·7)	93(5·7)	29(3·8)
AV Velocity (median [IQR])	4·40 [3·77, 5·00]	4·42 [3·79, 5·02]	4·44 [3·88, 5·00]	4·30 [3·30, 4·91]
AV Gradient (median [IQR])	48·00 [34·00, 65·00]	48·00 [34·00, 65·00]	48·40 [35·00, 64·00]	46·00 [28·25, 63·75]
PASP (median [IQR])	33·00[25·00,44·00]	32·00[24·00,44·00]	33·00[25·00,44·00]	35·00[30·00,45·00]
TAVR urgency (Selective TAVR) (%)	10642(98·6)	7785(98·5)	1955(98·9)	902(98·8)
TR level (%)				
None/Trace	1667(21·5)	1110(19·7)	301(20·9)	256(37·6)
Mild	4384(56·6)	3268(58·0)	832(57·8)	284(41·8)
Moderate	1296(16·7)	964(17·1)	221(15·4)	111(16·3)
Moderate-to-Severe	404(5·2)	290(5·1)	85(5·9)	29(4·3)
Severe	1(0·0)	1(0·0)	0(0·0)	0(0·0)
Access (%)				
Other access	38(0·4)	25(0·3)	5(0·3)	8(0·9)
TAo	8(0·1)	5(0·1)	2(0·1)	1(0·1)
Trans-subclavian	14(0·1)	11(0·1)	1(0·1)	2(0·2)
Trans-apical	747(6·9)	543(6·9)	130(6·6)	74(8·1)
Trans-carotid	107(1·0)	76(1·0)	17(0·9)	14(1·5)
Trans-femoral	9876(91·5)	7241(91·6)	1821(92·2)	814(89·2)
Circulatory support (%)	309(2·9)	239(3·0)	59(3·0)	11(1·2)
BNPratio (median [IQR])	3·27[1·00,9·65]	3·37[1·02,9·82]	3·43[1·03,10·00]	2·35[0·82,7·41]

Data were presented as number (percentage) or mean (interquartile range). Comparisons were performed using the Mann-Whitney U test or the Chi-square test. Abbreviation: BMI, body mass index; AV block, atrioventricular block; ALT, alanine aminotransferase; NYHA Class: New York heart association functiol classification; , not applicable; STS: Society of Thoracic Surgeons; LVOT, left ventricular outflow tract; LVIDd, left ventricular interl diameter at end-diastole; LVPWD, left ventricular posterior wall diameter; IVSd, interventricular septal diameter; LVEF, left ventricular ejection fraction; AR, aortic regurgitation; MR, mitral regurgitation; TR, tricuspid regurgitation; PASP, pulmory artery systolic pressure; BNP: b-type triuretic peptide.

3.2 Feature Selection

We performed feature selection and evaluation of multiple algorithms, ranking all variables according to their importance. The top 15 features identified by each algorithm were selected for further classifier modeling. (The results in each panel were generated by applying different algorithms: Double Input Symmetrical Relevance (DISR) (a), ANOVA-T-SCORE (b), LASSO (c), Random Forest (d), and XGBOOST (e)). Circulatory support emerged as the most important variable in both DISR and Anova t-score (Figs. 2a and 2b), while NT-pro BNP ranked first in RF and XGBoost (Figs. 2d and 2e). Additionally, variables such as renal function and height of coronary artery openings consistently ranked among the most important across all algorithms (Fig. 2).

Fig. 2 Feature Selection. Bar plots representing assessment of variable importance with 5 separate machine-learning algorithms, including a total of 35 variables. Blue bars indicate the top 15 most important features, while gray bars represent the remaining variables. Panel a was obtained by the DISR algorithm and indicates that the top 15 features are circulatory support, creatinine, TAVR urgency, ALT, NYHA class, AV block, left coronary height, LVEF, IVSd, BMI,

aortic angulation, LVIDd, hemoglobin, ascend aorta, and PASP. Panel b was obtained by the ANOVA-T-SCORE algorithm, Panel c by the LASSO algorithm, Panel d by the Random Forest algorithm and Panel e by the XGBOOST algorithm. The acronyms used are defined in the Table S2.

3.3 Model Development

Each feature-selection algorithm was combined with 15 different classifier algorithms, followed by an optimization process to identify the most appropriate number of features for each combination. The top 100 models, ranked by Area Under the ROC Curve (AUC), were retained and are displayed in Fig. 3a and Table S2. Among these, DISR + Support Vector Machine (SVM) model emerged as the top performer. Further refinement involved plotting the AUC variations of DISR with different classifiers across various feature subsets (Fig. 3b). Additional experiments with SVM as the constant classifier confirmed that DISR was the most effective feature-selection method for this algorithm (Fig. 3c). The final optimized model was DISR + SVM with five features, and a comprehensive evaluation of this model is provided in Figs. 3d and 3e and Table S3.

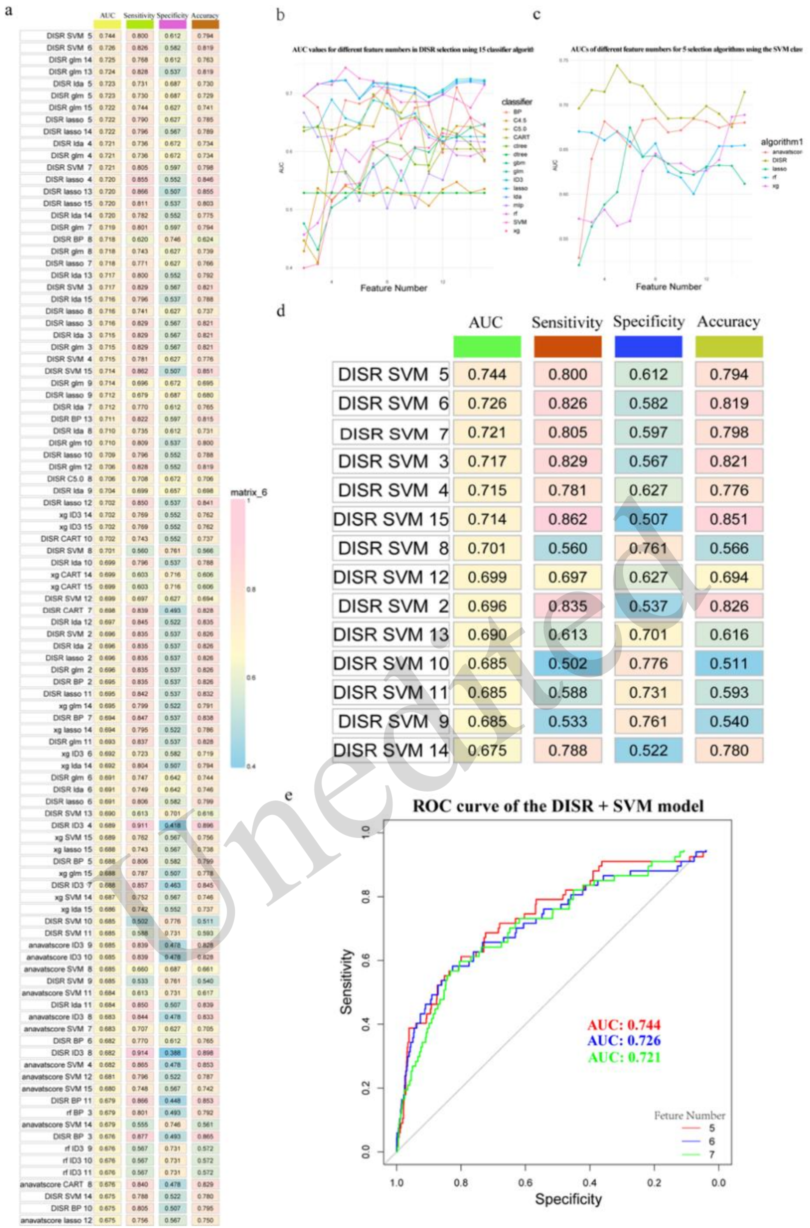

Fig. 3 Model performance. Each feature-selection algorithm was combined with 15 different classifier algorithms, and an optimization process was conducted to determine the most appropriate number of features for each combination. The top 100 models, ranked according to AUC, are shown in panel a. Figure b shows AUC values for different feature numbers in DISR selection, using 15 classifier algorithms. The best AUC value resulted from the DISR+SVM algorithm, with five feature numbers. Panel c shows the AUC values of different numbers for five selection algorithms, using SVM classifier algorithms. When the SVM classifier was fixed, the results confirmed that DISR was the optimal feature-selection method in combination with SVM. Panel d shows a comprehensive evaluation of the final model combination, detailing its AUC, accuracy, sensitivity, and specificity across different feature sets. Panel e shows the ROC curves for the DISR+SVM models with five, six, and seven features. AUC - area under the curve; DISR - Double Input Symmetrical Relevance; SVM - Support Vector Machine; glm - General Language Model; lda - Latent Dirichlet Allocation; lasso - Least Absolute Shrinkage and Selection Operator; BP - Backpropagation; C4.5, C5.0, and ID3 are types of decision tree algorithms;; CART -Classification and Regression Tree; ctree - Conditional Inference Tree; dtree - Decision Tree; gbm - Gradient Boosting Machine; mlp - Multilayer Perceptron; xg - Extreme Gradient Boosting.

3.4 Model Performance

The machine-learning model achieved an AUC of 0.74 (0.67, 0.81), outperforming the STS score, which had an AUC of 0.60 (0.53, 0.67) in internal validation (p = 0.004) (Fig. 4a). The machine-learning model also demonstrated superior accuracy, with a classification accuracy of 79%, compared to 77% for the STS score in the internal validation process. In external validation, the machine-learning model achieved an AUC of 0.69 (0.62, 0.76) and an accuracy of 83%, while the STS score showed an AUC of 0.57 (0.51, 0.63) (Fig. 4b) and an accuracy of 77%. Furthermore, the sensitivity and specificity of the machine-learning model were higher than those of the STS score in both internal and external validation (Internal validation: [0.80, 0.61] vs. [0.78, 0.30]; External validation: [0.84, 0.45] vs. [0.80, 0.28]). Additionally, the machine-learning model exhibited superior continuous Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI) over the STS score, with values of [0.66 (0.42, 0.90), 0.12 (0.08, 0.16)] in internal validation, and [0.70 (0.45, 0.94), 0.09 (0.05, 0.13)] in external validation (Table 2). The calibration plot for the machine-learning model is shown in Fig. S2. The first plot (on the left) corresponds to the internal validation set, and the second plot (on the right) corresponds to the external validation set. In addition to the visual assessment, we calculated quantitative calibration metrics. For the internal validation set, the calibration intercept was -3.3661, the calibration slope was 1.3877, and the Brier score was 0.1931. For the external validation set, the calibration intercept was -3.3699, the calibration slope was 1.034, and the Brier score was 0.1969, indicating agreement between predicted and observed outcomes. Decision-curve analysis further indicated that the machine-learning model significantly enhanced clinical decision-making (Fig. S3). A sensitivity analysis was conducted for both the machine-learning model and the STS score, including only cases with no missing data. Results indicated that the machine-learning model maintained strong performance in this subgroup (Table S4, Fig. S4).

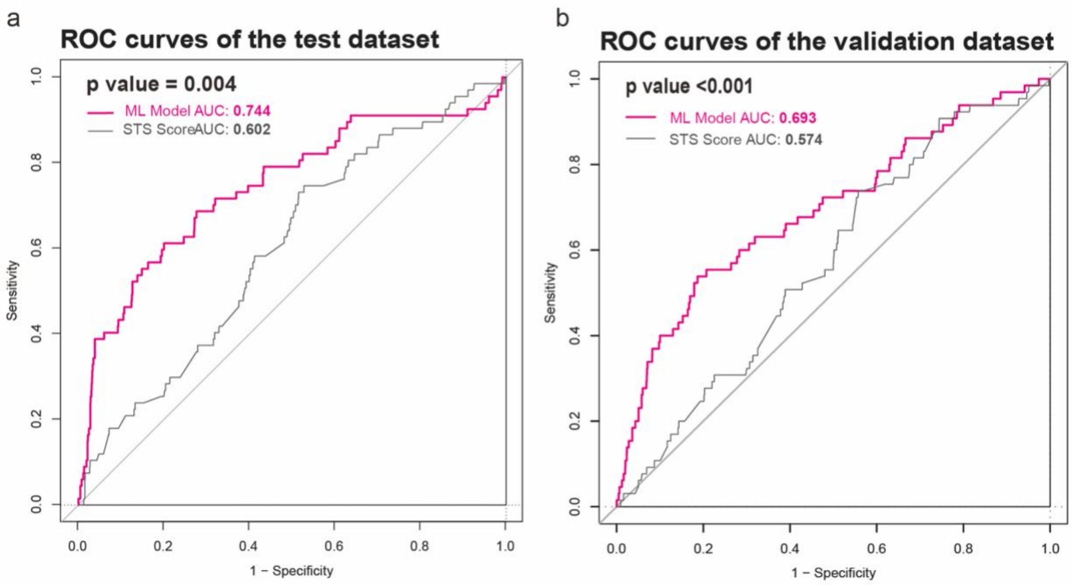

Fig. 4 Performance Comparison between the Machine-Learning Model and STS Score in Dataset and Validation

Dataset.

Table 2 Performance evaluation of the final model in the internal validation set and external validation sets

Model Index	Internal Validation		External Validation	
	ML-model	STS-score	ML-model	STS-score
AUC	0·74(0·67,0·81)	0·60(0·53,0·67)	0·69(0·62,0·76)	0·57(0·51,0·63)
Sensitivity	0·84(0·83,0·86)	0·78(0·77,0·80)	0·84(0·83,0·84)	0·79(0·77,0·81)
Specificity	0·55(0·43,0·67)	0·30(0·19,0·41)	0·45(0·33,0·57)	0·28(0·17,0·39)
Accuracy	0·83(0·82,0·85)	0·77(0·75,0·79)	0·83(0·81,0·85)	0·77(0·76,0·79)
Positive likelihood ratio	2·20(1·73,2·96)	1·39(1·20,1·65)	1·77(1·45,2·26)	1·34(1·17,1·61)
Negative likelihood ratio	0·29(0·23,0·38)	0·72(0·51,1·15)	0·35(0·27,0·48)	0·76(0·53,1·23)
Positive prediction	0·98(0·98,0·99)	0·97(0·96,0·98)	0·98(0·97,0·98)	0·97(0·96,0·98)
Negative prediction	0·10(0·07,0·13)	0·04(0·03,0·06)	0·08(0·05,0·11)	0·04(0·02,0·06)
F1 Score	0·91(0·90·0·92)	0·87(0·86,0·88)	0·91(0·90,0·92)	0·87(0·86,0·88)
NRI(Categorical)	0·31(0·17-0·45)	Ref.	0·22(0·06-0·39)	Ref.
NRI(Continuous)	0·66(0·42-0·90)	Ref.	0·70(0·45-0·94)	Ref.
IDI	0·12(0·08-0·16)	Ref.	0·09(0·05-0·13)	Ref.

3.5 Interpretability

SHAP analysis identified baseline alanine aminotransferase (ALT), baseline creatinine, baseline NYHA class, procedural circulatory support, and TAVR urgency as the most influential predictors of thirty-day mortality (Fig. 5, Fig. S5).

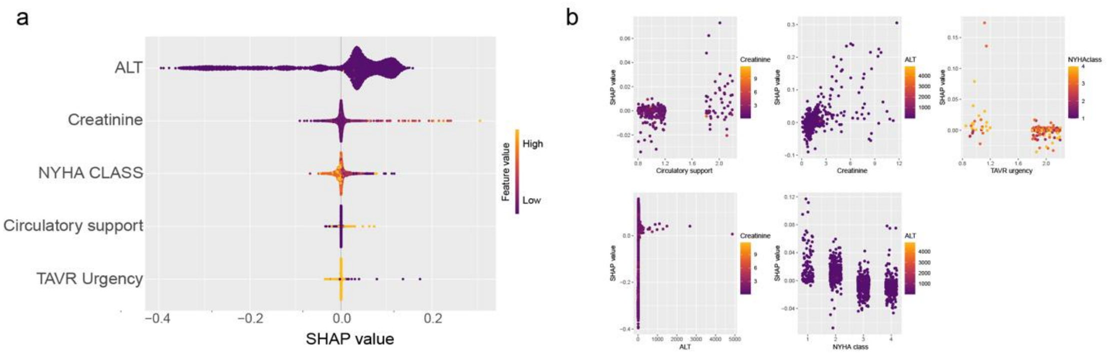

Fig. 5 SHapley Additive ExPlanations (SHAP) Summary Plot of Variable Importance for the Prediction Model. Predictive features are arranged based on their importance. Each dot represents one prediction result. SHAP values indicate the distribution of the prediction among the features; a positive value contributes to treatment success, while a negative value contributes to nonsuccess. SHAP analysis shows ALT, creatinine, NYHA class, circulatory support, and TAVR urgency as the most influential predictors of 30-day mortality. ALT - alanine aminotransferase; NYHA Class - New York Heart Hssociation Functional Classification.

4 Discussion

In this nationwide study, we developed a simple and effective machine-learning model for predicting short-term mortality in TAVR patients. By integrating multiple algorithms and feature-interpretation frameworks, we were able to create over 1,000 ML models for TAVR risk prediction. The final model not only delivers precise postoperative risk prediction but also identifies the key clinical parameters that influence outcomes in this population. This advance addresses a significant gap in preoperative TAVR assessment tools, providing clinicians with evidence-based, interpretable decision support that promotes the implementation of precision medicine in the management of structural heart disease.

This study reveals that the 30-day mortality risk for Chinese TAVR patients is 2.9%, consistent with recent reports from other national registries.(Arnold, et al., 2024; Rudolph et al., 2024) The current peri-procedural risk limits the potential for TAVR in younger patients. Therefore, individualized preoperative assessments and targeted screening for high-risk patients are essential. In the final model, baseline ALT levels, creatinine levels, NYHA class, TAVR urgency, and circulatory support are identifiable as key risk factors for thirty-day mortality. Circulatory support, in particular, reflects high intraoperative risk and serves as a comprehensive indicator of the patient's overall condition. Evaluating liver and renal function before TAVR is crucial, as postoperative recovery relies heavily on preserving multi-organ function.(Elbadawi et al., 2021; Witberg et al., 2021; Kolte et al., 2022) Recent studies also indicate that multi-organ dysfunction independently impacts outcomes.(Halavina et al., 2025) This finding underscores the importance of thorough preoperative and perioperative evaluation and appropriate management strategies.

The optimal machine-learning model consists of the top five features selected via the DISR method, modeled using an SVM classifier, which offers the best performance. In comparison to the traditional STS score, this model gives competitive results. The STS score is commonly used to predict postoperative risk in international TAVR patients.(Le Tourneau et al., 2010; Edwards et al., 2016; Arsalan et al., 2018) However, its applicability to Chinese patients is limited due to differences in data sources for model fitting, which diverge from those of the local patient population. This discrepancy introduces limitations when using the STS score to evaluate postoperative mortality risk in TAVR patients in China.(O'brien, et al., 2009; Jilaihawi, et al., 2015; Chen, et al., 2023) Other similar scoring systems have been utilized for TAVR risk assessment but still leave room for improvement.(Capodanno et al., 2014; Iung et al., 2014; Martin et al., 2018; Lantelme et al., 2019) Additionally, several studies have explored machine-learning algorithms to predict mortality risk in patients undergoing TAVR and have reported reasonable predictive performance.(Hernandez-Suarez et al., 2019; Gomes et al., 2021; Alhwiti et al., 2023; Leha et al., 2023) Our model demonstrates moderate discrimination, with performance that is numerically higher than that reported in previous studies of short-term mortality prediction after TAVR. Predicting early mortality in contemporary TAVR populations remains inherently challenging because event rates are relatively low and outcomes are influenced by complex interactions among patient comorbidities, anatomical characteristics, and procedural factors(Desai et al., 2021). In this context, moderate discrimination is commonly observed and is generally considered acceptable for clinical risk-prediction models. Importantly, the proposed model is intended to support risk stratification rather than provide deterministic predictions for individual patients, thereby assisting clinical decision-making in conjunction with comprehensive clinical evaluation. These findings suggest that the model may serve as a useful tool for pre-procedural risk stratification of postoperative mortality in patients undergoing TAVR.

This model can assist clinicians in identifying high-risk TAVR patients, enabling more personalized and effective management strategies. It also underscores the potential of machine learning to improve patient outcomes in clinical practice. Future research should prioritize validating machine-learning models across multi-ethnic cohorts and exploring their integration into real-world clinical workflows, such as registry-based risk-assessment tools or electronic decision-support systems. Such approaches could support pre-procedural risk stratification and multidisciplinary evaluation in patients undergoing TAVR. In addition, incorporating multi-omics data (e.g., genomics and proteomics) into future models may further enhance predictive performance.

Despite these promising results, our study has several limitations. First, while the data were derived from a multi-center cohort, and efforts were made to enroll all consecutive patients with aortic-valve disease, potential bias may still have been introduced due to the hospital-based nature of the cohort, as well as to patients who either refused consent or dropped out during follow-up. In addition, although SMOTE was applied only to the training set, oversampling may still have influenced probability calibration by altering the class distribution during model training. Therefore, further external validation is essential before implementing the model in clinical settings. Future studies should also focus on prospective validation and improving model performance by incorporating more diverse data, including multimodal data such as CT, echocardiographic imaging, and even genetic information.

5 Conclusion

In this study, a machine learning-based model was developed to predict 30-day mortality in Chinese TAVR patients. This model outperformed the traditional STS score and has promise as a valuable tool for enhancing patient care.

Data availability statement

The data used in this study contains sensitive patient information and cannot be shared publicly but may be made available upon request.

Acknowledgments

We appreciate very much for the participating hospitals to the registry and all the participants, also contributions made by Valve Center Alliance and China Heart House in the construction of Valve Centers. Thanks to all the experts of the expert committee of the Valve Center (https://www.valvecenter.org.cn/dynamic/dynamicDetail?id=27) and all Valve Center certification hospitals. This work was supported by the National Engineering Research Center of Science and TechnologyInformation STI Digital Health Laboratory (No. 2025STI136), the Zhejiang Provincial Science and Technology Program (No. 2024C03024)， the National Natural Science Foundation of China (No. U22A20267), and the National Natural Science Foundation of China (No. 82030014).

Author contributions

Qifeng ZHU, Jin LU, and Jiyuan LI contributed equally to this work. Xianbao LIU, and Jian'an WANG conceptualized the study. Qifeng ZHU, Feiyu WU, and Jiyuan LI conducted the data analysis, while Jin LU, Danqing YU, and Yihan PAN validated the findings. Qifeng ZHU and Jin LU drafted the manuscript. Xianbao LIU and Jian'an WANG provided overall supervision and project coordination. Qifeng ZHU, Qijing ZHOU, Chongzhou ZHENG, Daxin ZHOU and Wenzhi PAN contributed to critical review and intellectual input. Xianbao LIU, and Jian'an WANG revised the final manuscript. Xianbao LIU and Jian'an WANG are the guarantors of the study. Qifeng ZHU and Jin LU had full access to all study data and affirm its accuracy and integrity. Jian'an WANG made the final decision to submit the manuscript. The corresponding author confirms that all authors meet authorship criteria and that no eligible contributors have been omitted.

Compliance with ethics guidelines

Jian'an WANG is an Editorial Board Member for Joumal of Zhejiang University-SCIENCEB and was not involved in the editorial review or the decision to publish this article. All authors declare no financial or non-financial competing interests. The study protocol received approval from the Ethics Committee of the National Center for Cardiovascular Diseases of China (Approval No. 2017-968), with all participating centers granting their consent. Written informed consent was obtained from all eligible participants.

Declaration on the use of generative AI tools

In the preparation of this manuscript, ChatGPT was used solely for English language polishing and assistance with statistical code. No AI tools were employed for the generation of content, ideas, or interpretations, which are entirely the work of the authors.

References

Alhwiti T, Aldrugh S, Megahed FM, 2023. Predicting in-hospital mortality after transcatheter aortic valve replacement using administrative data and machine learning. *Sci Rep*, 13(1):10252. https://doi.org/10.1038/s41598-023-37358-9

Arnold SV, Manandhar P, Vemulapalli S, et al., 2024. Trends in transcatheter aortic valve replacement outcomes: Insights from the sts/acc tvt registry. *JAMA Cardiol*, 9(12):1115–1123. https://doi.org/10.1001/jamacardio.2024.3453

Arsalan M, Weferling M, Hecker F, et al., 2018. Tavi risk scoring using established versus new scoring systems: Role of the new sts/acc model. *EuroIntervention*, 13(13):1520–1526. https://doi.org/10.4244/eij-d-17-00421

Beohar N, Whisenant B, Kirtane AJ, et al., 2014. The relative performance characteristics of the logistic european system for cardiac operative risk evaluation score and the society of thoracic surgeons score in the placement of aortic transcatheter valves trial. *J Thorac Cardiovasc Surg*, 148(6):2830–2837.e2831. https://doi.org/10.1016/j.jtcvs.2014.04.006

Cai G, Huang F, Gao Y, et al., 2024. Artificial intelligence-based models enabling accurate diagnosis of ovarian cancer using laboratory tests in china: A multicentre, retrospective cohort study. *Lancet Digit Health*, 6(3):e176–e186. https://doi.org/10.1016/s2589-7500(23)00245-5

Capodanno D, Barbanti M, Tamburino C, et al., 2014. A simple risk tool (the observant score) for prediction of 30-day mortality after transcatheter aortic valve replacement. *Am J Cardiol*, 113(11):1851–1858. https://doi.org/10.1016/j.amjcard.2014.03.014

Carroll JD, Mack MJ, Vemulapalli S, et al., 2020. Sts-acc tvt registry of transcatheter aortic valve replacement. *J Am Coll Cardiol*, 76(21):2492–2516. https://doi.org/10.1016/j.jacc.2020.09.595

Chen J, Lyu L, Shen J, et al., 2023. Epidemiological study of calcified aortic valve stenosis in a chinese community population. *Postgrad Med J*, 99(1174):868–874. https://doi.org/10.1136/pmj-2022-141721

Collins GS, Moons KGM, Dhiman P, et al., 2024. Tripod+ai statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *Bmj*, 385:e078378. https://doi.org/10.1136/bmj-2023-078378

Deo RC, 2015. Machine learning in medicine. *Circulation*, 132(20):1920–1930. https://doi.org/10.1161/circulationaha.115.001593

Desai ND, O'brien SM, Cohen DJ, et al., 2021. Composite metric for benchmarking site performance in transcatheter aortic valve replacement: Results from the sts/acc tvt registry. *Circulation*, 144(3):186–194. https://doi.org/10.1161/circulationaha.120.051456

Edwards FH, Cohen DJ, O'brien SM, et al., 2016. Development and validation of a risk prediction model for in-hospital mortality after transcatheter aortic valve replacement. *JAMA Cardiol*, 1(1):46–52. https://doi.org/10.1001/jamacardio.2015.0326

Elbadawi A, Elzeneini M, Thakker R, et al., 2021. Transcatheter versus surgical aortic valve replacement in patients with combined chronic kidney and liver disease. *JACC Cardiovasc Interv*, 14(9):1047–1049. https://doi.org/10.1016/j.jcin.2021.03.009

Gomes B, Pilz M, Reich C, et al., 2021. Machine learning-based risk prediction of intrahospital clinical outcomes in patients undergoing tavi. *Clin Res Cardiol*, 110(3):343–356. https://doi.org/10.1007/s00392-020-01691-0

Halavina K, Koschatko S, Jantsch C, et al., 2025. Multiorgan dysfunction and its association with congestion and outcome in aortic stenosis treated with tavi. *JACC Adv*, 4(2):101544. https://doi.org/10.1016/j.jacadv.2024.101544

Hernandez-Suarez DF, Kim Y, Villablanca P, et al., 2019. Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. *JACC Cardiovasc Interv*, 12(14):1328–1338. https://doi.org/10.1016/j.jcin.2019.06.013

Hong N, Pan W, Zhou D, et al., 2022. The china heart valve center and national transcatheter valve therapeutics registry database. *Cardiology Plus*, 7(3):107–110.

Iung B, Vahanian A, 2012. Degenerative calcific aortic stenosis: A natural history. *Heart*, 98 Suppl 4:iv7–13. https://doi.org/10.1136/heartjnl-2012-302395

Iung B, Laouénan C, Himbert D, et al., 2014. Predictive factors of early mortality after transcatheter aortic valve implantation: Individual risk assessment using a simple score. *Heart*, 100(13):1016–1023. https://doi.org/10.1136/heartjnl-2013-305314

Jilaihawi H, Wu Y, Yang Y, et al., 2015. Morphological characteristics of severe aortic stenosis in china: Imaging corelab observations from the first chinese transcatheter aortic valve trial. *Catheter Cardiovasc Interv*, 85 Suppl 1:752–761. https://doi.org/10.1002/ccd.25863

Kolte D, Bhardwaj B, Lu M, et al., 2022. Association between early left ventricular ejection fraction improvement after transcatheter aortic valve replacement and 5-year clinical outcomes. *JAMA Cardiol*, 7(9):934–944. https://doi.org/10.1001/jamacardio.2022.2222

Kumar A, Sato K, Narayanswami J, et al., 2018. Current society of thoracic surgeons model reclassifies mortality risk in patients undergoing transcatheter aortic valve replacement. *Circ Cardiovasc Interv*, 11(9):e006664. https://doi.org/10.1161/circinterventions.118.006664

Lantelme P, Eltchaninoff H, Rabilloud M, et al., 2019. Development of a risk score based on aortic calcification to predict 1-year mortality after transcatheter aortic valve replacement. *JACC Cardiovasc Imaging*, 12(1):123–132. https://doi.org/10.1016/j.jcmg.2018.03.018

Le Tourneau T, Pellikka PA, Brown ML, et al., 2010. Clinical outcome of asymptomatic severe aortic stenosis with medical and surgical management: Importance of sts score at diagnosis. *Ann Thorac Surg*, 90(6):1876–1883. https://doi.org/10.1016/j.athoracsur.2010.07.070

Leha A, Huber C, Friede T, et al., 2023. Development and validation of explainable machine learning models for risk of mortality in transcatheter aortic valve implantation: Tavi risk machine scores. *Eur Heart J Digit Health*, 4(3):225–235. https://doi.org/10.1093/ehjdh/ztad021

Martin GP, Sperrin M, Ludman PF, et al., 2018. Novel united kingdom prognostic model for 30-day mortality following transcatheter aortic valve implantation. *Heart*, 104(13):1109–1116. https://doi.org/10.1136/heartjnl-2017-312489

Nashef SA, Roques F, Sharples LD, et al., 2012. Euroscore ii. *Eur J Cardiothorac Surg*, 41(4):734–744; discussion 744–735. https://doi.org/10.1093/ejcts/ezs043

O'brien SM, Shahian DM, Filardo G, et al., 2009. The society of thoracic surgeons 2008 cardiac surgery risk models: Part 2--isolated valve surgery. *Ann Thorac Surg*, 88(1 Suppl):S23–42. https://doi.org/10.1016/j.athoracsur.2009.05.056

Osnabrugge RL, Mylotte D, Head SJ, et al., 2013. Aortic stenosis in the elderly: Disease prevalence and number of candidates for transcatheter aortic valve replacement: A meta-analysis and modeling study. *J Am Coll Cardiol*, 62(11):1002–1012. https://doi.org/10.1016/j.jacc.2013.05.015

Otto CM, Nishimura RA, Bonow RO, et al., 2021. 2020 acc/aha guideline for the management of patients with valvular heart disease: A report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. *Circulation*, 143(5):e72–e227. https://doi.org/10.1161/cir.0000000000000923

Rudolph TK, Herrmann E, Bon D, et al., 2024. Comparison of contemporary transcatheter heart valve prostheses: Data from the german aortic valve registry (gary). *Clin Res Cardiol*, 113(1):75–85. https://doi.org/10.1007/s00392-023-02242-z

Sterne JA, White IR, Carlin JB, et al., 2009. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *Bmj*, 338:b2393. https://doi.org/10.1136/bmj.b2393

Structural Heart Disease Group of Chinese College of Cardiovascular Physician CHH, 2024. Chinese expert consensus on transfemoral transcatheter aortic valve replacement for pure aortic regurgitation (2023). *Cardiology Plus*, 9(3):217–226. https://doi.org/10.1097/cp9.0000000000000090

Vahanian A, Beyersdorf F, Praz F, et al., 2022. 2021 esc/eacts guidelines for the management of valvular heart disease. *Eur Heart J*, 43(7):561–632. https://doi.org/10.1093/eurheartj/ehab395

Witberg G, Steinmetz T, Landes U, et al., 2021. Change in kidney function and 2-year mortality after transcatheter aortic valve replacement. *JAMA Netw Open*, 4(3):e213296. https://doi.org/10.1001/jamanetworkopen.2021.3296

Xu H, Liu Q, Cao K, et al., 2022. Distribution, characteristics, and management of older patients with valvular heart disease in china: China-dvd study. *JACC Asia*, 2(3):354–365. https://doi.org/10.1016/j.jacasi.2021.11.013

Supplementary information

Tables S1–S4, Figs. S1–S5