POLISH HEART JOURNAL

Kardiologia Polska
The Official Peer-reviewed Journal of the Polish Cardiac Society since 1957

Online first
This is a provisional PDF only. Copyedited and fully formatted version will be made available soon

ISSN 0022–9032
e-ISSN 1897–4279

Application of machine learning in predicting postoperative arrhythmia following transcatheter closure of perimembranous ventricular septal defects

Authors: Lintao Yan, Yuan Meng, Hongjie Sun, Xinlei Liu, Bo Han

Article type: Original article

Received: June 11, 2024

Accepted: November 7, 2024

Early publication date: November 25, 2024

Application of machine learning in predicting postoperative arrhythmia following transcatheter closure of perimembranous ventricular septal defects

Short title: ML predicts arrhythmia post-pmVSD closure

Lintao Yan[1], Yuan Meng[2], Hongjie Sun[3], Xinlei Liu[4], Bo Han[1,2]

[1]Department of Pediatric Cardiology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China

[2]Department of Pediatric Cardiology, Shandong Provincial Hospital Affiliated to Shandong 1st Medical University, Jinan, Shandong, China

Correspondence to:

Bo Han, MD,

Department of Pediatric Cardiology,

Shandong Provincial Hospital Affiliated to Shandong 1st Medical University,

Jingwu Road 324, 250021 Jinan, Shandong, China,

phone: +86 13 60 531 39 27,

e-mail: hanbo35@163.com, 198657000056@email.sdu.edu.cn

WHAT'S NEW?

This study integrates multi-dimensional data, including physical examinations, medical history, electrocardiogram, echocardiograms, etc. for comprehensive analysis. Using advanced machine learning algorithms, such as logistic regression, extreme gradient boosting, support vector machines, and random forest, we constructed a predictive model. The developed risk assessment tool accurately predicts arrhythmias, facilitating precise forecasting. This data-driven tool helps clinicians make informed pre- and post-operative decisions, thereby improving patient outcomes. Interdisciplinary collaboration ensures the model's broad applicability and generalizability.

ABSTRACT

Background: Arrhythmia is a frequent complication following transcatheter device closure of perimembranous ventricular septal defects (pmVSD). However, there is currently a lack of a convenient predictive tool for postoperative arrhythmia.

Aims: This research aims to use machine learning algorithms to predict the risk of postoperative arrhythmia in patients with pmVSD.

Methods: A retrospective study was conducted on 1384 children with pmVSD who underwent successful transcatheter device closure at a single-center hospital from March 2002 to March 2024. Subjects were divided into a training set (n = 970) and a validation set (n = 414) in a 7:3 ratio. Four machine learning methods — SVM, LR, RF, and XGBOOST — develop models for predicting postoperative arrhythmia using preoperative and intraoperative baseline information with clinical significance, as well as relevant content mentioned in previously published journals. Models performance were evaluated using area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, negative predictive value, and positive predictive value. The optimal model was used to create a nomogram, further calibrated with calibration curves.

Results: In the prediction of postoperative arrhythmias, the LR model outperformed the XGBOOST, SVM, and RF models, achieving an AUC of 0.863 (95% CI, 0.827–0.900). Consequently, we utilized the LR model to construct a nomogram based on 5 variables: weight, procedure time, defect diameter, pre-interventional arrhythmia, and the diameter difference between the occluder and defect exceeding 2 mm. The calibration curves illustrated a strong agreement between the actual and predicted outcomes.

Conclusions: A machine learning model accurately predicts postoperative arrhythmias, aiding in risk stratification of patients with pmVSD and guiding clinical decisions.

Key words: arrhythmia, machine learning, perimembranous ventricular septal defect, prediction model, transcatheter closure,

INTRODUCTION

Ventricular septal defect (VSD) is a congenital cardiac condition where tissue abnormalities in the ventricular septum result in abnormal interventricular blood flow. About 40% of all congenital heart illnesses are related to this condition, which is one of the most prevalent. The most prevalent subtype is the perimembranous ventricular septal defects (pmVSD), accounting for 70% of cases [1, 2]. Transcatheter interventional occlusion for pmVSD has become a preferred option over traditional surgical procedures, thanks to advancements in occlusion devices and interventional techniques. This method is less invasive, has a quicker recovery time, and is more cost-effective compared to open-heart surgery [3, 4]. However, patients undergoing surgery are still at risk of postoperative complications, with an incidence ranging from 20% to 30%. Among these complications, arrhythmia is a common occurrence, with conduction blocks such as bundle branch block and atrioventricular block being the primary types. Severe blocks can be life-threatening, necessitating the removal of the blocker, repair of the VSD, and installation of a pacemaker [2, 3]. Therefore, accurately predicting the occurrence of postoperative arrhythmias is crucial to assist healthcare providers in assessing the risk of developing arrhythmias in different scenarios and guiding clinical decision-making towards precision medicine. Unfortunately, existing studies lack predictive models for assessing the risk of early postoperative arrhythmias following transcatheter occlusion in patients with pmVSD. With recent advancements in artificial intelligence and the increased use of machine learning (ML) in the medical field, this study aims to develop and validate an interpretable clinical predictive model using multiple machine learning methods and visualize its outcomes [5–7].

MATERIAL AND METHODS

Study population and design

This retrospective study analyzed data from 1384 patients with pmVSD who underwent successful transcatheter device closure at the Department of Pediatric Cardiology, Provincial Hospital of Shandong First Medical University between March 2002 and March 2024. Inclusion criteria: age ≥2 years or weight ≥10 kg, confirmed diagnosis of hemodynamically significant pmVSD (e.g., cardiomegaly; left atrial enlargement; left ventricular volume overload on chest radiographs), defect detected by transthoracic echocardiography at the 9–12 o’clock position in the short-axis parasternal view, and pulmonary artery systolic blood pressure <70 mm Hg on transthoracic echocardiography. Exclusion criteria comprised non-membranous VSD, concomitant interventional closure procedures, and patients with missing crucial clinical data. Patient data, including vital signs and preoperative, intraoperative, and postoperative details, were sourced from the Platform for Epidemiological Investigation and Precision Treatment of Childhood Heart Disease, established by the Department of Pediatric Cardiology at the Provincial Hospital of the First Medical University of Shandong, China (http://www.pedhd.cn/). The study received approval from the local institutional ethics committee (Ethics Committee of Shandong First Medical University Hospital), and written informed consent was obtained from all patients’ guardians. Data anonymization and confidentiality were ensured, and the study adhered to the principles of the Declaration of Helsinki.

The study assessed the potential factors for the model's construction based on clinical significance, scientific understanding, and predictors found in previously published publications. [8]. The variables evaluated in this study included preoperative clinical assessment data such as height, weight, body surface area, and age, along with risk factors identified from previous studies and selected based on clinical significance. These risk factors encompassed Pre-interventional arrhythmia (cardiac rhythm abnormalities present before the intervention), ventricular septal defect diameter (measured via ultrasound during intraoperative angiography), presence of membranous aneurysm (an abnormal bulging or aneurysmal structure formed due to structural abnormality or weakness in the membranous portion of the ventricular septum), and distance from the aortic valve to the defect (measured between the defect margin and the aortic valve, with a threshold of 3 mm). Additional procedural parameters included the type and diameter of occluder devices (comprising Amplatzer Duct Occluder II, eccentric occluder, symmetric occluder, and thin-waist occluder), the diameter difference between the occluder and ventricular septal defect (Ddov, with a threshold of 2 mm), procedure time, and first-attempt success (defined as successful device placement without requiring replacement with an occluder of a different type or diameter) [9–18].

Preoperative information, including gender, age, height, weight, and body surface area, was measured and calculated by the preoperative nurse. Preoperative arrhythmia was determined by multiple cardiologists through electrocardiogram examinations. Intraoperative information, such as defect diameter, tumor of the membranous part of the interventricular septum, and the distance of the aortic valve to the defect, was observed during intraoperative imaging. The procedure time, diameter and type of the occluder, the diameter difference between the occluder and ventricular septal defect, and the success of the first attempt at occlusion were recorded by the operating surgeon.

Machine learning algorithms

XGBoost is an integrated learning algorithm based on decision trees, implemented through a gradient boosting framework, which is able to efficiently deal with non-linear relationships and high-dimensional data, and is particularly suitable for large-scale datasets. Its ability to handle missing values and unbalanced data makes it excellent in complex clinical data. Logistic regression is a classical statistical method for estimating the probability of an event occurring and is suitable for binary classification problems. Because of its computational efficiency and ease of interpretation, logistic regression is widely used in clinical settings where explicit interpretation of model outputs is required. SVM performs classification by finding the optimal segmentation hyperplane in a high-dimensional feature space, which is suitable for handling small samples and high-dimensional data. It performs well in classification tasks with high-dimensional features such as gene expression data and can provide highly accurate classification results with small samples. Random forest improves model accuracy and stability by constructing multiple decision trees and combining their predictions. Its feature importance scoring function helps to identify key variables, which is particularly useful for classification and feature selection for complex conditions [7, 19–22].

Ethics

Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Statistical analysis

To evaluate the normality of continuous variables, the Shapiro–Wilk test was employed. The test results indicate that the continuous variables in this study deviate from a normal distribution, thus we provided the median and interquartile range (IQR) for continuous variables, and we used the Mann-Whitney U test for comparisons between groups. Frequencies and percentages were used to summarize categorical variables. The chi-square test or Fisher exact test, if applicable, was used to compare categorical variables between groups. A two-tailed *P*-value of less than 0.05 was deemed statistically significant.

In order to prevent multicollinearity, stepwise regression was carried out by root Akaike information criterion and evaluated by Variance Inflation Factor. All samples were randomly selected, with 70% serving as the training set and the remaining 30% as the validation set. Logistic regression was performed using the glm function to fit the data. Family parameters were set as binomial. Since the tune.svm function in the e1071 package allows for cross-validation optimization of tuning parameters and kernel functions, we utilize it to construct linear SVM models. builds a linear SVM model using the e1071 package. operates by means of cross-validation. The random forest model was constructed using the random forest function found in the random forest package. There are 292 trees built, and the least mean of the squared residuals is used to identify the particular ideal tree. The xgboost package was used to build the model, and cross-validation was used to fine-tune the parameters. Make use of 10-fold cross-validation optimization to validate every model. The pROC and caret packages may be employed to calculate the area under the receiver operating characteristic (ROC) curve (AUC), as well as accuracy, sensitivity, specificity, negative predictive value, and positive predictive value in order to evaluate predictive performance. The effectiveness of a classification model can be determined by comparing its AUC, which indicates how near the model is to 1. Using the rms software, a nomogram was created for the best model based on the findings of the multivariable analysis. Each regression coefficient in the multivariable logistic regression is scaled and converted into a scale ranging from 0 to 100 to create the nomogram. A score of 100 is given to the variable whose effect has the highest beta coefficient (in absolute value). A total score is obtained by adding the scores of the independent variables, and this value is subsequently translated into a projected probability. The calibration curve was used to evaluate the relationship between the probability of actual observations and the predicted probability, which in a well-calibrated model should fall on a 45-degree diagonal. The consistency index was used to evaluate the nomograms' ability to discriminate. Internal validation was carried out utilizing bootstrapping (1000 resamplings) to prevent possible overfitting [23]. The flowchart is shown in Figure 1, In all analyses, P <0.05 was considered statistically significant. All data analyses R version 4.33 were performed.

RESULTS

Patient characteristics

In this study, 1550 patients underwent transcatheter closure of VSD, with 111 patients having non-pmVSD. Additionally, 39 patients had other transcatheter device closures, and 16 patients had severely incomplete clinical data. Ultimately, 1384 patients met the inclusion criteria, with 475 experiencing postoperative arrhythmias. The median age of the patients was 3.75 years, and 50.3% were female. The data were divided into a training set (n = 970) and a validation set (n = 414) in a 7:3 ratio for the modeling process. There were no significant baseline characteristic differences between the training and validation sets. Table 1 presents the univariate analysis of the total patient dataset and the baseline characteristics of the training and validation sets.

Machine learning algorithms and comparison prediction model performance

Four machine learning algorithms were used to construct our prediction models: XGBOOST, LR, SVM, and RF. We then plotted the ROC curves and assessed each model's predictive performance in terms of AUC, sensitivity, specificity, precision, positive predictive value, and negative predictive value (Table 2 and Figure 2A–B). Five variables were ultimately chosen to be included in the logistic regression model . On multivariable analysis, weight, procedure time, defect diameter, pre-interventional arrhythmia, and the diameter difference between the occluder and defect exceeding 2 mm were independently associated with postoperative arrhythmia (Table 3), with Figure 3 displaying the relevance ratings of the variables obtained from the random forest regressor. Weight, procedure time, defect diameter, The diameter difference between the occluder and ventricular septal defect >2 mm (Ddov2), and Pre-interventional arrhythmia (pre-ARR) were the factors included in the model. The training set's AUC values for XGBOOST, LR, SVM, and RF were 0.955, 0.822, 0.920, and 1, respectively; the validation set's AUC values for the same variables were 0.849, 0.863, 0.859, and 0.848. In the end, the order of AUC values indicated that the LR model was the best model.

Development and validation of an ARR-predicting nomogram

The following variables were identified as independent risk factors for the development of ARR in patients with pmVSD following interventional occlusion: body weight, procedure time, defect diameter, diameter difference between the occluder and ventricular septal defect >2 mm, and pre-interventional arrhythmia. The aforementioned five variables were incorporated into the predictive model and utilized to construct the nomogram (Figure 4). In comparison to complex logistic regression formulas, nomograms are relatively straightforward to comprehend and offer greater clinical utility. Based on the score of each independent variable, a score can be derived by projecting vertically to the top score axis. The corresponding total score is then located below, and the sum of the scores of each variable is projected onto the risk axis of postoperative ARR to predict the incidence of ARR in patients after pmVSD. The higher the total score, the higher the risk of ARR. The nomogram was calibrated using calibration curves. The calibration curve's findings demonstrated that there was little deviation between the ideal and real curves. The prediction model has good accuracy because the actual prediction curves (B = 1000 resamples, mean absolute error of training set [A] = 0.012, mean absolute error of validation set [B] = 0.02) were consistent with the ideal curves. The prediction model has some predictive value following internal validation (Figure 5A–B). The nomogram's C-statistic is 0.822, indicating a high predictive value.

DISCUSSION

Technology advancements have made transcatheter device closure the more popular treatment option for patients with pmVSD, but postoperative arrhythmias remain the most common and serious side effect. Atrioventricular block and complete left bundle branch block, in particular, can be extremely dangerous and even life-threatening, necessitating the installation of a pacemaker. According to past research, recovery from transcatheter blocker arrhythmias is more challenging the earlier they occur [24]. Consequently, a straightforward and intelligible instrument for gauging the probability of postoperative arrhythmia occurrences is of paramount importance for clinicians' surgical decision-making and early prompt intervention in postoperative high-risk children. By providing a more accurate assessment of the patient's overall condition prior to operation, it helps guide the selection of appropriate surgical techniques and optimal occluder size and type, thereby reducing the risk of postoperative arrhythmias. It also facilitates tailored post-operative monitoring plans, lifestyle adjustments and psychological interventions for high-risk patients. Early detection and management of arrhythmias based on predictive models can prevent disease progression and improve patient recovery and long-term outcomes. Therefore, emphasizing the importance of postoperative arrhythmia prediction in clinical practice is critical to optimizing treatment strategies, improving patient care, and improving overall prognosis.

In this study, four machine learning algorithms were employed and validated to construct a prediction model to predict the probability of postoperative arrhythmia in patients with pmVSD who underwent transcatheter device closure. The performance of the ML algorithms was evaluated, and the prediction model built by logistic regression was determined to be the optimal model. This model can assist clinicians in their decision-making.

The model we developed incorporated several preoperative intraoperative risk factors as the most important predictors of postoperative ARR, including pre-interventional arrhythmia, body weight, defect diameter, The diameter difference between the occluder and ventricular septal defect, and procedure time. Our study found that pre-interventional arrhythmia plays a crucial predictive role. It has been shown that pre-interventional arrhythmia accounts for over 70% of postoperative arrhythmia cases in children after VSD transcatheter device closure. Pre-interventional arrhythmia is associated with hemodynamic abnormalities that lead to cardiac conduction block. If these hemodynamic abnormalities are not corrected after the procedure, the combination of surgical stimulation and increased conduction bundle compression can easily trigger arrhythmias [18]. The cardiovascular system of children with lower body weight is immature, making them more sensitive to surgical stress and increasing the probability of postoperative arrhythmia [25, 26]. The size of the defect is also a crucial factor; a larger VSD significantly impacts the heart's anatomy and electrical activity. Studies show that larger defect sealing may lead to a more significant local inflammatory response and scar tissue formation, affecting the stability of the heart’s electrical activity [27, 28]. In pmVSD, the anatomic alignment of the atrioventricular (AV) conduction system varies. Typically, the His bundle travels along the posterior-inferior border of the defect, with its perforating branches only 2–4 mm from the edge. The left and right bundle branches, along with other branches, are sometimes wrapped in residual fibrous tissue at the edge of the defect [29]. Therefore, when the diameter difference between the occluder and the ventricular septal defect is greater than 2 mm, meaning the occluder is too large relative to the defect, it compresses the AV junction area. This leads to oedema of the surrounding myocardial tissue, causing arrhythmias. This phenomenon is understandable from both clinical practice and anatomical perspectives. The longer the procedure, the more cardiac tissue manipulation is involved, which may increase local inflammation and affect the heart’s electrical conduction system. Prolonged exposure to catheters and occlusion devices can mechanically irritate cardiac tissues, especially conduction pathways, thereby increasing the likelihood of arrhythmic events. This is consistent with many studies’ findings [2, 30].

More than 3 decades have passed since the introduction of transcatheter ventricular septal defect sealing, and many studies on postoperative arrhythmias have been conducted. However, various biases have arisen, likely due to inconsistent or overly detailed classifications of arrhythmias, small sample sizes, incomplete data, and confounding non-independent risk factors. Consequently, many mixed results have been obtained. Moreover, the research literature on applying ML algorithms to predict arrhythmias after pmVSD is sparse. To our knowledge, this is the first study to develop a predictive model using ML algorithms to guide postoperative arrhythmia risk assessment from general clinical information. Utilizing advances in computer science to incorporate clinical parameters may be more beneficial for risk prediction than relying on a single clinical parameter. The prediction model in this study establishes a visual nomogram that can intuitively predict the risk of developing postoperative arrhythmias. Clinical staff can use the nomogram to calculate postoperative risk probabilities, aiding in intervention and clinical decision-making. Calibration curves showed the model has good accuracy. Therefore, developing a clinical predictive modeling tool would facilitate timely interventions to prevent or mitigate postoperative arrhythmias. However, there are limitations to this study. First, the nature of retrospective studies may lead to selection bias. Second, our ML algorithmic model is somewhat limited to specific mechanisms, which may restrict its generalization. This aspect needs further validation in real scenarios. Additionally, our current clinical database is a single-center database only for Chinese patients, limiting its predictive value. Finally, since our database includes only pediatric patients, the predictive validity for adult patients with post-pmVSD arrhythmias is unknown. Therefore, our next step is to establish a large, heterogeneous, multicenter, multinational, multi-age cohort to improve, homogenize, and validate the current clinical risk prediction models.

CONCLUSION

By analyzing clinical data from our database, we developed and evaluated machine learning algorithms to predict postoperative ARR risk in pmVSD patients using available preoperative and intraoperative variables. The logistic regression model performed the best. We also constructed an easy-to-use and internally validated nomogram. Clinicians can use this nomogram to predict the risk of postoperative ARR in pmVSD patients, enabling accurate surgical decisions and effective risk communication for better prognoses. The nomogram requires a larger sample size for further external validation.

Article information

Conflict of interest: None declared.

Funding: This work was supported by the Natural Science Foundation of Shandong Province [ZR2023MH052].

REFERENCES

1. Penny DJ, Vick GW. Ventricular septal defect. Lancet. 2011; 377(9771): 1103–1112, doi: 10.1016/s0140-6736(10)61339-6, indexed in Pubmed: 21349577.
2. Jiang D, Zhang S, Zhang Y, et al. Predictors and long-term outcomes of heart block after transcatheter device closure of perimembranous ventricular septal defect. Front Cardiovasc Med. 2022; 9: 1041852, doi: 10.3389/fcvm.2022.1041852, indexed in Pubmed: 36386314.
3. Weryński P, Skorek P, Wójcik A, et al. Recent achievements in transcatheter closure of ventricular septal defects: a systematic review of literature and a meta-analysis. Kardiol Pol. 2021; 79(2): 161–169, doi: 10.33963/KP.15708, indexed in Pubmed: 33331744.
4. Saurav A, Kaushik M, Mahesh Alla V, et al. Comparison of percutaneous device closure versus surgical closure of peri-membranous ventricular septal defects: A systematic review and meta-analysis. Catheter Cardiovasc Interv. 2015; 86(6): 1048–1056, doi: 10.1002/ccd.26097, indexed in Pubmed: 26257085.
5. Merkin A, Krishnamurthi R, Medvedev ON. Machine learning, artificial intelligence and the prediction of dementia. Curr Opin Psychiatry. 2022; 35(2): 123–129, doi: 10.1097/YCO.0000000000000768, indexed in Pubmed: 34861656.
6. de Jong G, Aquarius R, Sanaan B, et al. Prediction models in aneurysmal subarachnoid hemorrhage: Forecasting clinical outcome with artificial intelligence. Neurosurgery. 2021; 88(5): E427–E434, doi: 10.1093/neuros/nyaa581, indexed in Pubmed: 33548918.
7. Chen Y, Zheng Y, Lip GYH. Leveraging machine learning for enhanced mortality risk prediction in atrial fibrillation: A step towards precision medicine? Pol Heart J. 2024; 82(10): 921–923, doi: 10.33963/v.phj.102455, indexed in Pubmed: 39240922.

8. Kim Y, Margonis GA, Prescott JD, et al. Nomograms to predict recurrence-free and overall survival after curative resection of adrenocortical carcinoma. JAMA Surg. 2016; 151(4): 365–373, doi: 10.1001/jamasurg.2015.4516, indexed in Pubmed: 26676603.
9. Holzer R, de Giovanni Jo, Walsh KP, et al. Transcatheter closure of perimembranous ventricular septal defects using the Amplatzer membranous VSD occluder: Immediate and midterm results of an international registry. Catheter Cardiovasc Interv. 2006; 68(4): 620–628, doi: 10.1002/ccd.20659, indexed in Pubmed: 16969878.
10. Bergmann M, Germann CP, Nordmeyer J, et al. Short- and long-term outcome after interventional VSD closure: A single-center experience in pediatric and adult patients. Pediatr Cardiol. 2021; 42(1): 78–88, doi: 10.1007/s00246-020-02456-2, indexed in Pubmed: 33009919.
11. Xian SJ, Wei G, Ai ZQ, et al. Risk factors for arrhythmia early after transcatheter closure of perimembranous ventricular septal defects. Chin J Pediatr. 2005; 43(10): 767–771.
12. Wang L, Cao S, Li J, et al. Transcatheter closure of congenital perimembranous ventricular septal defect in children using symmetric occluders: An 8-year multiinstitutional experience. Ann Thorac Surg. 2012; 94(2): 592–598, doi: 10.1016/j.athoracsur.2012.03.067, indexed in Pubmed: 22626754.
13. Zuo J, Xie J, Yi W, et al. Results of transcatheter closure of perimembranous ventricular septal defect. Am J Cardiol. 2010; 106(7): 1034–1037, doi: 10.1016/j.amjcard.2010.05.040, indexed in Pubmed: 20854970.
14. Zhao LJ, Han Bo, Zhang JJ, et al. Postprocedural outcomes and risk factors for arrhythmias following transcatheter closure of congenital perimembranous ventricular septal defect: A single-center retrospective study. Chin Med J (Engl). 2017; 130(5): 516–521, doi: 10.4103/0366-6999.200551, indexed in Pubmed: 28229981.
15. Wu Z, Yang P, Xiang P, et al. Left anterior fascicular block after transcatheter closure of ventricular septal defect in children. Front Cardiovasc Med. 2021; 8: 609531, doi: 10.3389/fcvm.2021.609531, indexed in Pubmed: 34179122.
16. Lei YQ, Lin WH, Lin SH, et al. Influence of percutaneous catheter intervention for

congenital perimembranous ventricular septal defects in children on the cardiac conduction system and associated risk factors: a meta-analysis. J Cardiothorac Surg. 2022; 17(1): 19, doi: 10.1186/s13019-022-01751-8, indexed in Pubmed: 35172864.

17. Tang C, Shao S, Zhou K, et al. Complete left bundle-branch block after transcatheter closure of perimembranous ventricular septal defect using Amplatzer duct occluder II. J Am Heart Assoc. 2022; 11(7): e022651, doi: 10.1161/JAHA.121.022651, indexed in Pubmed: 35350872.
18. Li Y, Hua Y, Fang J, et al. Identification of risk factors for arrhythmia post transcatheter closure of perimembranous ventricular septal defect. J Invasive Cardiol. 2015; 27(8): E158–E166, indexed in Pubmed: 26232018.
19. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 785–794, doi: 10.1145/2939672.2939785.
20. Cherkassky V, Ma Y. Another look at statistical learning theory and regularization. Neural Netw. 2009; 22(7): 958–969, doi: 10.1016/j.neunet.2009.04.005, indexed in Pubmed: 19443179.
21. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20(3): 273–297, doi: 10.1007/bf00994018.
22. Breiman L. Random forests. Machine Learning. 2001; 45(1): 5–32.
23. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. JAMA. 2017; 318(14): 1377–1384, doi: 10.1001/jama.2017.12126, indexed in Pubmed: 29049590.
24. Yang R, Kong XQ, Sheng YH, et al. Risk factors and outcomes of post-procedure heart blocks after transcatheter device closure of perimembranous ventricular septal defect. JACC Cardiovasc Interv. 2012; 5(4): 422–427, doi: 10.1016/j.jcin.2012.01.015, indexed in Pubmed: 22516400.
25. Tanghöj G, Liuba P, Sjöberg G, et al. Risk factors for adverse events within one year after atrial septal closure in children: A retrospective follow-up study. Cardiol Young. 2020; 30(3): 303–312, doi: 10.1017/S1047951119002919, indexed in

Pubmed: 31847927.

26. Kramer H, Körfer R. Early postoperative arrhythmias after cardiac operation in patients with congenital heart disease. Pediatr Cardiol. 2005; 26(1): 10–15.
27. Kovach JR, Brown ML. Postoperative arrhythmias after congenital heart surgery. J Thorac Dis. 2013; 5(S6): S526–S533.
28. El-Said HG. Transcatheter closure of congenital ventricular septal defects. Cath Cardiovasc Interv. 2003; 60(1): 107–113.
29. Terada R, Warren S, Lu JT, et al. Ablation of Nkx2-5 at mid-embryonic stage results in premature lethality and cardiac malformation. Cardiovasc Res. 2011; 91(2): 289–299, doi: 10.1093/cvr/cvr037, indexed in Pubmed: 21285290.
30. Gatzoulis MA, Balaji S, Webber SA, et al. Risk factors for arrhythmia and sudden cardiac death late after repair of tetralogy of Fallot: A multicentre study. Lancet. 2000; 356(9234): 975–981, doi: 10.1016/S0140-6736(00)02714-8, indexed in Pubmed: 11041398.

Table 1. Perioperative statistical data of participants

Characteristic	**Total (n = 1384)**		**Training (n = 970)**		**Validation (n = 414)**		
	NO-ARR	**ARR**	**NO-ARR**	**ARR**	**NO-ARR**	**ARR**	***P*-value**
	n = 909	**n = 475**	**n = 637**	**n = 333**	**n = 272**	**n = 142**	
Sex, n (%)							0.130
F	471 (51.8%)	225 (47.4%)	329 (51.6%)	161 (48.3%)	142 (52.2%)	64 (45.1%)	

M	438 (48.2%)	250 (52.6 %)	308 (48.4 %)	172 (51.7%)	130 (47.8 %)	78 (54.9 %)	
Weight, kg, median (IQR)	16.0 (14.0–21.0)	15.0 (13.0–19.0)	16.0 (14.0–20.5)	15.0 (13.5–19.0)	16.5 (14.0 – 22.0)	15.0 (13.0 – 18.0)	<0.0 01
Height, cm, median (IQR)	103 (96.5–114)	100 (95.0–110)	102 (96.3–113)	100 (95.3–110)	104 (96.9 –115)	99.0 (95.0 –106)	<0.0 01
Bsa, m^2, median (IQR)	0.68 (0.60–0.81)	0.63 (0.59–0.76)	0.67 (0.60–0.80)	0.64 (0.59–0.77)	0.69 (0.60 – 0.83)	0.63 (0.58 – 0.72)	<0.0 01
Age, years, median (IQR)	3.83 (2.92–5.50)	3.50 (2.75–5.00)	3.83 (2.92–5.42)	3.50 (2.75–5.25)	3.96 (3.00 – 5.71)	3.42 (2.75 – 4.42)	0.00 1
Tom, n (%)							0.05 1
NO	240 (26.4%)	102 (21.5 %)	179 (28.1 %)	72 (21.6%)	61 (22.4 %)	30 (21.1 %)	
YES	669(73.6 %)	373 (78.5 %)	458 (71.9 %)	261 (78.4%)	211 (77.6 %)	112 (78.9 %)	
Procedure time, min, median	60.0 (55.0–80.0)	70.0 (60.0 –110)	60.0 (55.0–80.0)	70.0 (60.0–110)	60.0 (55.0 –	77.5 (60.0 –110)	<0.0 01

(IQR)					85.0)		
Type, n (%)							0.023
Ado2	135 (14.9%)	63 (13.3%)	99 (15.5%)	47 (14.1%)	36 (13.2%)	16 (11.3%)	
Ecc	83 (9.13%)	52 (10.9%)	58 (9.11%)	38 (11.4%)	25 (9.19%)	14 (9.86%)	
Sym	641 (70.5%)	315 (66.3%)	447 (70.2%)	218 (65.5%)	194 (71.3%)	97 (68.3%)	
Tw	50 (5.50%)	45 (9.47%)	33 (5.18%)	30 (9.01%)	17 (6.25%)	15 (10.6%)	
Dev, mm, median (IQR)	6.00 (5.00–7.00)	6.00 (5.00–8.00)	6.00 (5.00–7.00)	6.00 (5.00–8.00)	6.00 (5.00–7.00)	7.00 (6.00–8.00)	<0.001
Fs, n (%)							<0.001
NO	97 (10.7%)	100 (21.1%)	71 (11.1%)	69 (20.7%)	26 (9.56%)	31 (21.8%)	
YES	812 (89.3%)	375 (78.9%)	566 (88.9%)	264 (79.3%)	246 (90.4%)	111 (78.2%)	
Dd, mm, median	3.10 (2.30–	4.00 (2.90–	3.10 (2.30–	4.00 (2.80–	3.05 (2.46	4.45 (3.00	<0.0

(IQR)	4.23)	5.90)	4.25)	5.80)	– 4.14)	– 6.00)	01
Davd, n (%)							1.000
>3 mm	432 (47.5%)	226 (47.6%)	296 (46.5%)	156 (46.8%)	136 (50.0%)	70 (49.3%)	
≤3 mm	477 (52.5%)	249 (52.4%)	341 (53.5%)	177 (53.2%)	136 (50.0%)	72 (50.7%)	
Ddov, n (%)							<0.001
≤2 mm	366 (40.3%)	102 (21.5%)	263 (41.3%)	74 (22.2%)	103 (37.9%)	28 (19.7%)	
>2 mm	543 (59.7%)	373 (78.5%)	374 (58.7%)	259 (77.8%)	169 (62.1%)	114 (80.3%)	
Pre-Arr, n (%)							<0.001
NO	903 (99.3%)	305 (64.2%)	635 (99.7%)	220 (66.1%)	268 (98.5%)	85 (59.9%)	
YES	6 (0.66%)	170 (35.8%)	2 (0.31%)	113 (33.9%)	4 (1.47%)	57 (40.1%)	

Abbreviations: ado2, Amplatzer Duct Occluder II; Davd, the distance of the aortic valve to defect; Dd, defect diameter; Ddov, the diameter difference between the occluder and ventricular septal defect; Ecc, eccentric occlude; Fs, success of the first attempt; NO-ARR, no postoperative arrhythmia; Pre-Arr, pre-interventional arrhythmia; Sym, symmetric occlude; Tom, Tumor of membranous part of interventricular septum;Tw, thin-waist occlude

Table 2. Predictive performance of four machine learning models in the training and validation cohorts

TRAINING SET						
Model	**AUC**	**Sen**	**Spe**	**Acc**	**PPV**	**NPV**
XGB	0.955	0.925	0.834	0.865	0.743	0.955
LR	0.822	0.619	0.885	0.794	0.738	0.816
SVM	0.92	0.822	0.903	0.875	0.815	0.907
RF	1	1	1	1	1	1

VALIDATION SET						
Model	**AUC**	**Sen**	**Spe**	**Acc**	**PPV**	**NPV**
XGB	0.849	0.734	0.816	0.788	0.677	0.854
LR	0.863	0.683	0.868	0.804	0.729	0.84
SVM	0.859	0.776	0.827	0.81	0.703	0.875
RF	0.848	0.685	0.857	0.798	0.715	0.838

Abbreviations: NPV, negative predictive value; PPV, positive predictive value

Table 3. Multivariable logistic regression analysis on the training set

	[ALL] **n = 970**	**NO-ARR** **n = 637**	**ARR** **n = 333**	***P*-value**
Weight, kg	16.0 (14.0–20.0)	16.0 (14.0–20.5)	15.0 (13.5–19.0)	0.009
Procedure time,	70.0 (60.0–90.0)	60.0 (55.0–80.0)	70.0 (60.0–110)	<0.001

min				
Dd, mm	3.50 (2.50–4.90)	3.10 (2.30–4.25)	4.00 (2.80–5.80)	<0.001
Ddov:				<0.001
≤2mm	337 (34.7%)	263 (41.3%)	74 (22.2%)	
>2mm	633 (65.3%)	374 (58.7%)	259 (77.8%)	
Pre-Arr:				<0.001
NO	855 (88.1%)	635 (99.7%)	220 (66.1%)	
YES	115 (11.9%)	2 (0.31%)	113 (33.9%)	

Abbreviations: NO-ARR, no postoperative arrhythmia; Dd, defect diameter; Ddov, the diameter difference between the occluder and ventricular septal defect; Pre-Arr, pre-interventional arrhythmia

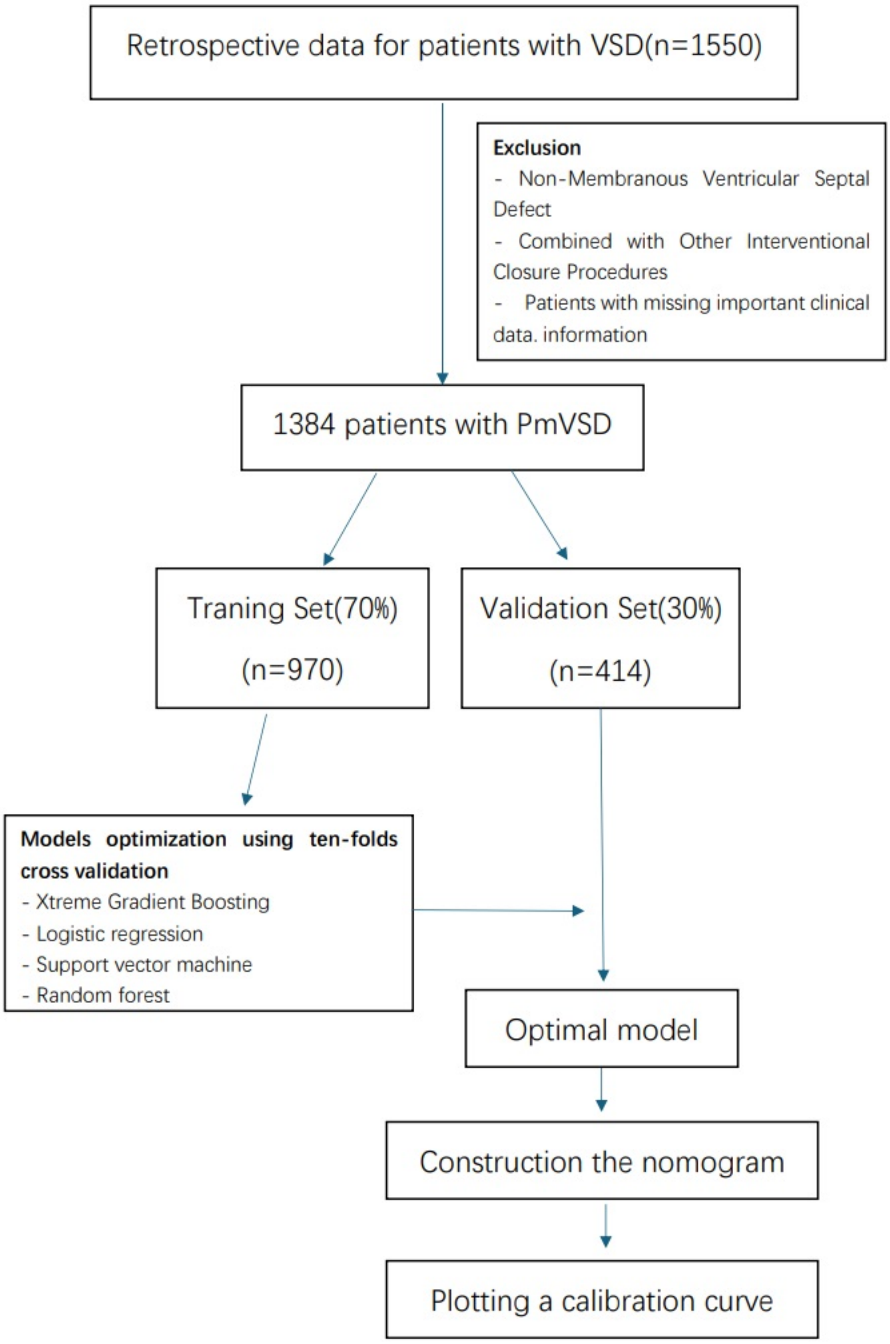


Figure 1. The flowchart of the study

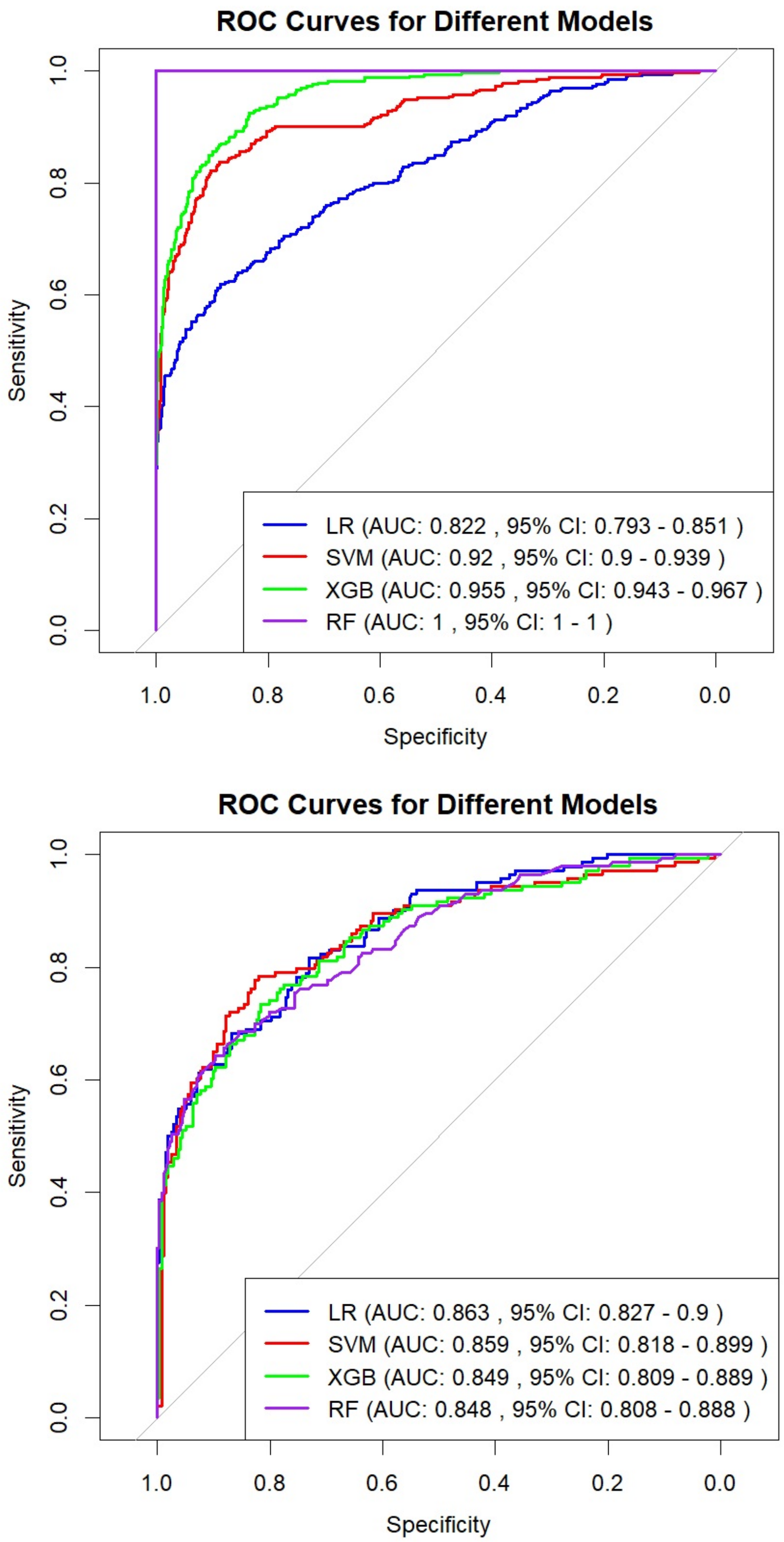


Figure 2. The ROC curves and AUC values with 95% CI for the four machine learning prediction models on the training and validation sets

Abbreviations: AUC, area under the curve; CI, confidence interval; ROC, receiver operating characteristic

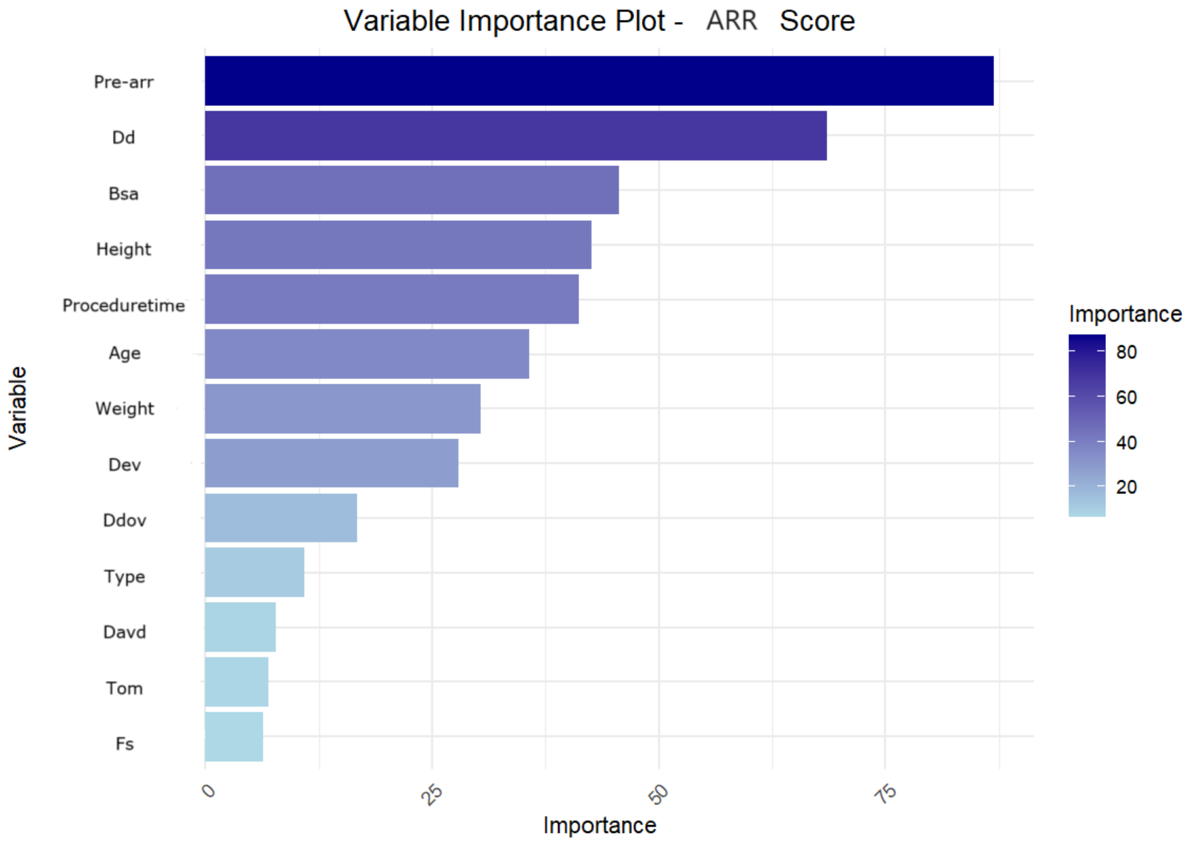


Figure 3. The important features derived from the random forest regressor

Abbreviations: Pre-Arr, pre-interventional arrhythmia; Dd, defect diameter; Dev, device diameter; Ddov, the diameter difference between the occluder and ventricular septal defect; Davd, the distance of the aortic valve to defect; Tom, tumor of membranous part of interventricular septum; Fs, success of the first attempt

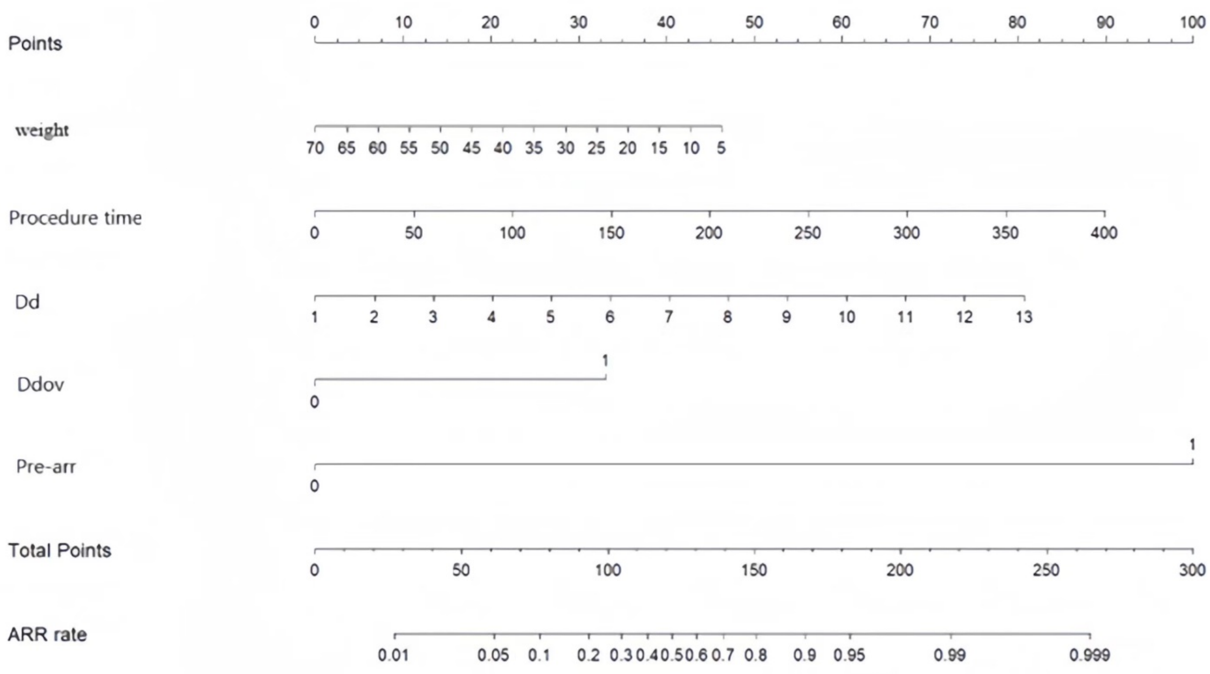
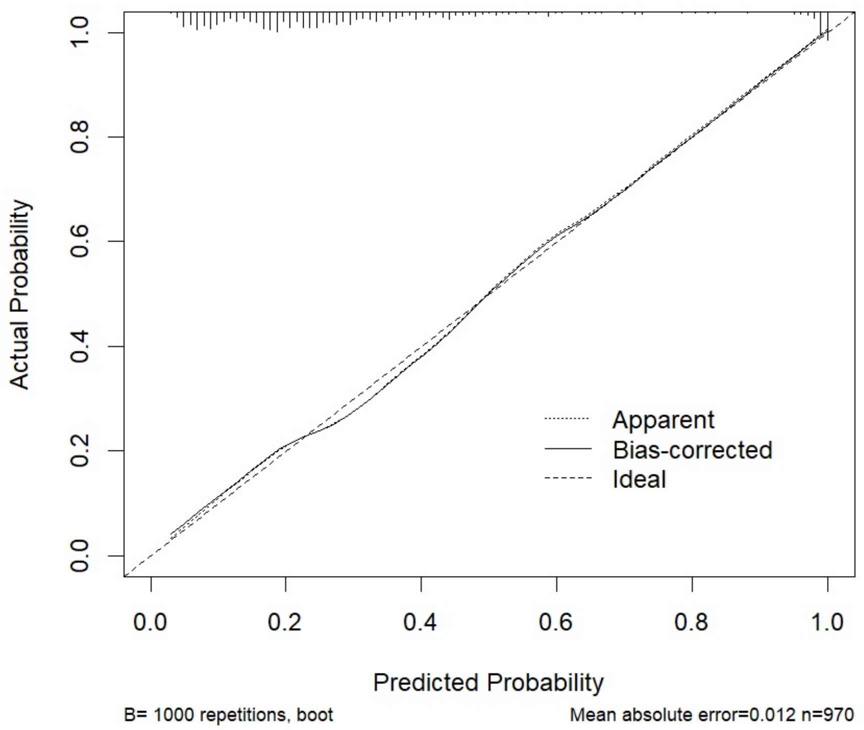
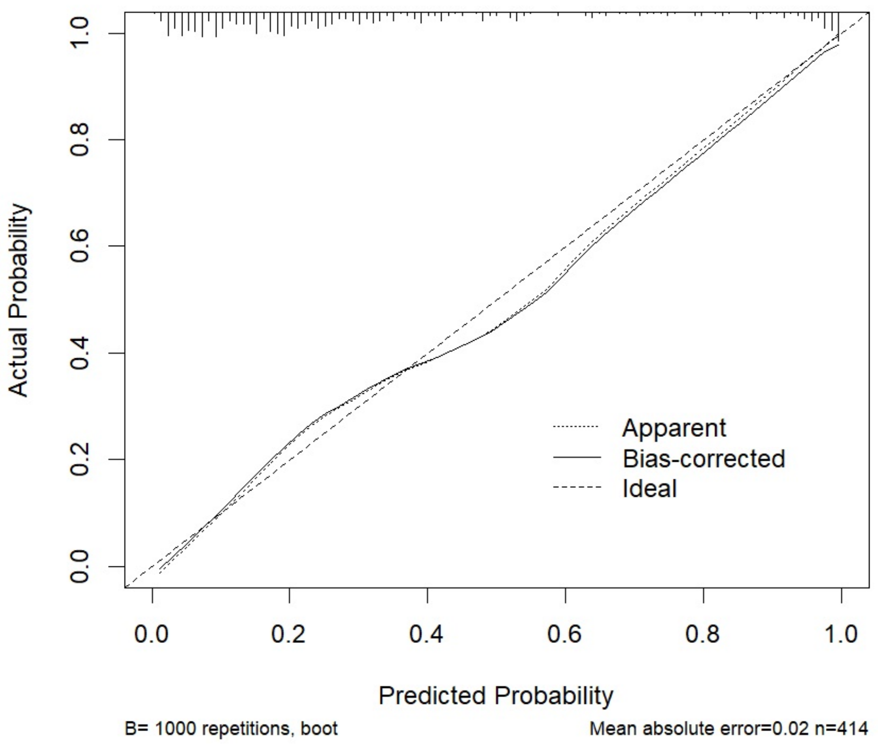


Figure 4. Nomogram used for predicting ARR after interventional occlusion in pmVSD patients. Logistic regression algorithm was used to establish nomogram. The total points is calculated as the sum of the individual scores of each of the five variables included in the nomogram



1.0
0.8
0.6
0.4
0.2
0.0
Actual Probability
0.0
0.2
0.4
0.6
0.8
1.0
Predicted Probability
Apparent
Bias-corrected
Ideal
B= 1000 repetitions, boot
Mean absolute error=0.012 n=970



1.0
0.8
0.6
0.4
0.2
0.0
Actual Probability
0.0
0.2
0.4
0.6
0.8
1.0
Predicted Probability
Apparent
Bias-corrected
Ideal
B= 1000 repetitions, boot
Mean absolute error=0.02 n=414

Figure 5. Calibration curve of the nomogram for the training set (**A**) and the validation set (**B**). The logistic regression algorithm was used to establish a nomogram. The X-axis represents the overall predicted probability of ARR after transcatheter device closure in pmVSD patients. and the Y-axis represents the actual probability. Model calibration is indicated by the degree of fitting of the curve and the diagonal