Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Application of Machine Learning Models Integrating Clinical and Echocardiography in the Prediction of Mean Pulmonary Artery Pressure Grading

Xinpeng Dai
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Qiumeng Xi
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Jiayi He
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Rui Fan
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Xinyuan Zhang
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Dichen Guo
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Juanni Gong
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Suqiao Yang
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Yuanhua Yang
Beijing Chaoyang Hospital Affiliated to Capital Medical University
Yidan Li
yidan_li@163.com

Beijing Chaoyang Hospital Affiliated to Capital Medical University

Research Article

Keywords: Pulmonary hypertension, Mean pulmonary artery pressure, Echocardiography, Machine learning, Model interpretability

Posted Date: March 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8424355/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

As a progressive cardiopulmonary disorder, pulmonary hypertension (PH) necessitates precise assessment of mean pulmonary arterial pressure (mPAP) for clinical staging, treatment planning, and prognostic evaluation.

Methods

We retrospectively included patients who underwent right heart catheterization (RHC) at our institution between January 2017 and October 2025. The cohort was divided temporally into a training cohort (January 2017 to December 2023) and a validation cohort (January 2024 to October 2025). Echocardiographic parameters and clinical data were collected. A four-category label was constructed based on mPAP grading (0–20 mmHg, 21–35 mmHg, 36–45 mmHg, > 45 mmHg). Key features were selected using Lasso combined with the Boruta method. The Synthetic Minority Over-sampling Technique (SMOTE) balanced the training cohort sample distribution. Ultimately, eight machine learning (ML) models were constructed and their performance evaluated. Model performance was assessed using area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, F1 score, and Brier score. Feature importance for predictive models was interpreted using SHapley Additive exPlanations (SHAP) values.

Results

A total of 495 patients were included in model construction. Six features were selected from 29 variables for model training: 6-Minute Walk Distance (6MWD), Eccentricity Index (EI), Left Ventricular Diameter (LVD), Right Ventricular Diameter (RVD), Tricuspid Annular Plane Systolic Excursion/Pulmonary Artery Systolic Pressure (TAPSE/PASP), and PASP. Among all ML models, the Naive Bayes model achieved the highest classification accuracy, with an AUC of 0.886, accuracy of 0.736, Brier score of 0.106, and F1 score of 0.736. Its AUC within the training cohort reached 0.894. Furthermore, the mean AUC values for different mPAP classifications were 0.994, 0.878, 0.779, and 0.892, respectively. SHAP value analysis confirmed that TAPSE/PASP was the primary predictive feature for mPAP classification, followed by PASP and EI. These three features demonstrated consistent performance across all subgroups.

Conclusions

The non-invasive predictive model developed in this study provides a reliable tool for the precise classification of mPAP in PH patients, thereby assisting clinicians in reducing reliance on invasive RHC.

Background

Pulmonary hypertension (PH) is a severe cardiopulmonary disorder characterized by progressive elevation of pulmonary vascular resistance [1]. Without timely and effective treatment, the condition progressively deteriorates, leading to right heart failure and even death, posing a grave threat to patients' health and life [2, 3]. Currently, right heart catheterization (RHC) serves as the clinical gold standard for diagnosing PH and accurately measuring mean pulmonary artery pressure (mPAP) [4]. Through RHC, clinicians can directly obtain pulmonary artery pressure data, providing crucial evidence for disease confirmation and severity assessment. However, RHC is an invasive procedure. Not only is the operation complex, demanding high standards of medical technique and equipment, but it also carries certain operational risks, potentially triggering complications such as hemorrhage, infection, and arrhythmia [5].

Transthoracic echocardiography (TTE) is a widely employed, convenient, and cost-effective non-invasive diagnostic modality in clinical practice. It plays a crucial role in the diagnosis and assessment of PH [6]. TTE provides a clear visualization of cardiac and great vessel structure and function, enabling the acquisition of key parameters such as right ventricular dimensions, pulmonary artery diameter, and tricuspid regurgitation velocity. Existing research indicates a strong correlation between these echocardiographic parameters and pulmonary artery pressure [7, 8]. Yang et al. developed composite indices combining the tricuspid regurgitation gradient with the right pulmonary artery diameter and the main pulmonary artery diameter with the right pulmonary artery diameter, achieving positive predictive values of 95.2% and 95.4%, respectively, for predicting mPAP ≥ 20 mmHg [9]. However, traditional analysis methods based on a single or a few echocardiographic indicators have limitations. They struggle to comprehensively and accurately capture the complex information embedded in echocardiographic data and the non-linear relationships between parameters. This leads to errors in the precise grading and assessment of PH, failing to meet the urgent clinical demand for accurate diagnosis and personalized treatment.

The rapid advancements in information technology have led to the increased application of machine learning (ML) techniques in the medical field, which has resulted in significant potential being demonstrated [10, 11]. Unlike traditional data analysis methods, ML can automatically mine latent patterns and regularities from large volumes of complex data by constructing data-driven models [12]. In the context of medical prediction tasks, ML has demonstrated its capacity to integrate multi-dimensional medical data to establish accurate prediction models, thereby providing a scientific basis for disease diagnosis, treatment, and prognostic evaluation [13, 14]. For instance, in the context of tumor diagnosis, ML algorithms are capable of analyzing pathological images, genetic data, and other information, with a view to improving the accuracy of early diagnoses [15, 16]. In the field of echocardiography, ML has emerged as a promising tool for predicting the onset risk and disease progression probability of cardiovascular diseases. By analyzing patients' cardiac structural parameters, hemodynamic indicators, and myocardial motion characteristics, among others, ML assists clinicians in formulating personalized diagnoses and treatment plans [17–19].

In the field of PH diagnosis and staging, ML also holds vast application potential. Hirata et al. employed a logistic regression model based on elastic network regularization methods utilizing ultrasound and clinical parameters, achieving area under the curve values of 0.789, 0.766, and 0.742 for diagnosing patients with normal PH, precapillary PH, and postcapillary PH, respectively [20]. Zhao et al. developed and validated a multimodal deep learning model demonstrating superior specificity and negative predictive value compared to conventional TTE in PH detection, with robustness across different patient subgroups [21]. Although ML has yielded certain results in PH, most current research focuses solely on whether patients have PH or not, with relatively few studies addressing the prediction of the four PH classifications [22, 23]. Therefore, developing a four-classification prediction model for PH based on ultrasound data holds significant theoretical importance and practical application value for enhancing the diagnostic accuracy and clinical management of PH.

Methods

Study Subjects and Data Sources

This study retrospectively analyzed patients who underwent RHC at our hospital between January 2017 and October 2025. Based on previously established research criteria [24, 25], patients were divided into four groups according to the mPAP measured by RHC: normal group (mPAP ≤ 20 mmHg), mild elevation group (20 mmHg < mPAP ≤ 35 mmHg), moderate elevation group (35 mmHg < mPAP ≤ 45 mmHg), and severe elevation group (mPAP > 45 mmHg). The exclusion criteria were set as follows: (1) no TTE performed; (2) interval between TTE and RHC exceeding 7 days; (3) missing rate of core variables required for analysis > 20%. Finally, a total of 495 patients were included in the study. To verify the generalizability of the model, a time-series splitting strategy was adopted to partition the dataset: the training cohort included 374 patients from January 2017 to December 2023, and the independent validation cohort included 121 patients from January 2024 to October 2025. The study protocol was approved by the Ethics Committee of Beijing Chaoyang Hospital. Since this study is retrospective, the requirement for informed consent was waived.

Right Heart Catheterization

RHC was performed using a Swan-Ganz catheter. Measurements were completed with the patient in the supine position at end-expiratory. Recorded hemodynamic parameters included mPAP, central venous pressure (CVP), mean pulmonary capillary wedge pressure (PCWP), systolic pulmonary artery pressure (sPAP), diastolic pulmonary artery pressure (dPAP), cardiac output (CO), and pulmonary vascular resistance (PVR).

Echocardiography Examination

All TTE examinations were performed by specialists with echocardiography expertise using a commercial ultrasound system (EPIQ7C, Philips Healthcare, Massachusetts, USA) equipped with an X5-1 phased-array transducer. The following echocardiographic parameters were separately collected: Left Ventricular Ejection Fraction (LVEF), Systolic Pulmonary Artery Pressure (PASP, calculated from tricuspid regurgitation velocity), Eccentricity Index (EI), Left Ventricular Diameter (LVD), Right Ventricular Diameter (RVD), Main pulmonary artery diameter (MPA), right pulmonary artery diameter (RPA), left pulmonary artery diameter (LPA), right ventricular end-diastolic area (RVEDA), right ventricular end-systolic area (RVESA), fractional area change (FAC), tricuspid annular systolic excursion (TAPSE), TAPSE/PASP, myocardial performance index (MPI), peak systolic velocity of the tricuspid annulus (S'), peak early diastolic velocity of the tricuspid annulus (E'), and peak late diastolic velocity of the tricuspid annulus (A'). All clinical data were collected from the electronic health record system, namely: disease subgroup, sex, age, WHO functional class (I-IV), body mass index (BMI), body surface area (BSA), heart rate (HR), SpO_2 (oxygen saturation), N-terminal pro-B-type natriuretic peptide (NT-proBNP), and 6-minute walk distance (6MWD). For missing values, K-nearest neighbor interpolation was employed. Figure 1 illustrates the entire workflow of this study.

Feature Selection

To further refine the features, a feature selection framework that combines the Least Absolute Shrinkage and Selection Operator (LASSO) regression method with the Boruta algorithm was employed [25]. First, all features were standardized using Z-scores. Subsequently, unbiased feature selection was performed using the Boruta algorithm, with a random forest regressor of maximum depth 5 serving as the base model. Shadow features were generated, and statistically significant predictive features were identified through permutation testing. Concurrently, a λ logarithmic grid was constructed via Lasso regression. The optimal regularization parameter was selected through 10-fold cross-validation repeated three times, utilizing L1 regularization to compress redundant feature coefficients to zero. Finally, the intersection of features selected by both methods was extracted as the consensus core features.

ML Model Construction

To address the issue of imbalanced distribution in the training cohort's original samples and mitigate the risk of overfitting, we employed the Synthetic Minority Over-sampling Technique (SMOTE) to resolve data imbalance. By synthesizing minority class samples to augment their representation within the dataset, we achieved sample equilibrium, thereby enhancing the model's recognition capability for the minority class. Subsequently, we employed eight ML algorithms to construct models: Logistic Regression (LR), Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes (NB), Extreme Gradient Boosting (XGB), and Decision Tree (DT). We employed a rigorous data partitioning methodology. Specifically, the entire dataset was first divided into a training cohort and an independent validation cohort. Within the training cohort, this was further subdivided into an internal training cohort and an internal testing cohort. By implementing 10-fold cross-validation, the entire training cohort was randomly partitioned into ten equal segments. Each iteration selected nine segments as the internal training cohort, with the remaining segment serving as the internal testing cohort for

model parameter optimization. This process is repeated ten times to ensure each subset is used as the test cohort once. The optimal combination of hyperparameters is determined through these ten cross-validation iterations, with the model's optimal parameters detailed in Supplementary Table 1. Finally, we evaluate the model's generalization capability by testing its performance on the independent validation cohort. Throughout this process, strict independence between the training and validation cohorts is maintained to ensure data isolation and prevent data leakage risks.

Model Evaluation and Feature Interpretation

The primary metric for evaluating model performance is area under the receiver operating characteristic curve (AUC). To estimate the 95% confidence interval for AUC, we employed a non-parametric bootstrap method with 1000 resamples. Additional performance metrics include accuracy, recall, F1 score, and Brier score. Category-specific metrics comprise sensitivity, specificity, and precision for each classification level. To quantify the contribution of each variable to model predictions, feature importance was assessed. SHapley Additive exPlanations (SHAP) were employed to analyze the contribution of each input variable to the model output. Global interpretation was achieved by plotting a bar chart of the mean absolute SHAP values, providing an intuitive representation of the overall importance of each feature. Subsequently, SHAP histograms for each category were used to elucidate the importance of individual features in predicting different categories.

Statistical Analysis

The distribution of continuous variables was assessed using the Kolmogorov-Smirnov test. Variables meeting normality assumptions were analyzed using independent samples t-tests, with results presented as mean ± standard deviation. For non-normally distributed variables, the Mann-Whitney U test was employed, with results displayed as median (interquartile range) in the format M(Q1, Q3). Categorical variables were described by frequency (percentage), with intergroup comparisons performed using the chi-square test or Fisher's exact test. Statistical significance was set at $p < 0.05$. All statistical analyses and model construction were completed using Python software (version 3.12.9).

Results

Initially, we retrospectively reviewed 675 patients who underwent RHC examinations. Following application of exclusion criteria, 180 patients were excluded, with details presented in Fig. 2. Ultimately, our study cohort comprised 495 patients (mean age 56 years [range 45–65]; 235 males [47.5%]). Table 1 summarizes baseline characteristics for patients in the training and validation cohorts. The training cohort (n = 374) and validation cohort (n = 121) were broadly comparable overall, but TTE parameters revealed significant differences in RVD, LVD, RVEDA, E', and A' between the two cohorts.In the clinical baseline data, significant differences existed between the two cohorts in terms of BSA and 6MWD, whereas no differences were observed in RHC results. Intra-cohort analyses of the training and validation cohorts revealed that the vast majority of features exhibited significant changes with increasing mPAP classification (Supplementary Table 2). Notably, higher mPAP classifications were associated with significantly elevated NT-proBNP levels and markedly reduced 6MWD. Among TTE parameters, higher mPAP grades were associated with significant increases in PASP, EI, pulmonary artery width, and right ventricular area, alongside significant decreases in TAPSE and S'.

This study incorporated 30 clinical and echocardiographic variables. Following Lasso regression and Boruta screening, six features demonstrated strong associations with mPAP grading: 6MWD, EI, LVD, RVD, PASP, and TAPSE/PASP. Although PASP and the TAPSE/PASP ratio are correlated, PASP directly reflects pulmonary arterial pressure load, whereas the TAPSE/PASP ratio accounts for right ventricular-pulmonary artery coupling status. Consequently, both features were retained.

Table 1
Baseline characteristics of patients in the training and validation cohorts

Variables	Training cohort	Validation cohort	*p*-value
Demographics			
Age, years	56.00(45.00, 66.00)	56.00(47.00, 63.00)	0.986
Sex			
Male, %	171 (45.7%)	64 (52.9%)	0.057
Female, %	203 (54.3%)	57 (47.1%)	
BMI, kg/m^2	23.99 ± 3.77	24.22 ± 3.32	0.537
BSA, m^2	1.69(1.58, 1.82)	1.76(1.63, 1.86)	**0.016**
Clinical parameters			
NT-proBNP, pg/ml	276.10(82.50, 1103.98)	360.00(86.10, 1020.00)	0.748
6MWD, m	399.00(300.00, 464.75)	378.00(240.00, 450.00)	**0.040**
HR, bpm	79.51 ± 13.03	79.11 ± 11.24	0.763
sBP, mmHg	119.00(108.00, 130.75)	121.00(110.00, 135.00)	0.079
dBP, mmHg	74.17 ± 11.28	76.31 ± 10.53	0.067
SpO_2	94.00(90.00, 96.00)	93.00(90.00, 95.00)	0.050
Subgroup			
1	130 (34.8%)	31 (25.6%)	0.110
2	4 (1.1%)	0 (0.0%)	
3	1 (0.3%)	0 (0.0%)	
4	232 (62%)	86 (71.1%)	
5	7 (1.9%)	4 (3.3%)	
WHO FC			
1	54 (14.4%)	17 (14.0%)	0.075
2	208 (55.6%)	78 (64.5%)	
3	83 (22.2%)	24 (19.8%)	
4	29 (7.8%)	2 (1.7%)	
Echocardiography			
LVEF, %	63.00(53.08, 68.00)	61.0278(43.63, 67.02)	0.052
PASP, mmHg	71.69 ± 27.70	75.62 ± 28.86	0.179
EI	1.24(1.06, 1.42)	1.30(1.10, 1.42)	0.123
LVD, mm	37.20(33.00, 44.00)	36.00(32.00, 40.70)	**0.043**
RVD, mm	42.59 ± 8.12	44.51 ± 8.32	**0.025**
MPA, mm	31.70(28.00, 34.26)	31.00(28.60, 34.00)	0.908
RPA, mm	22.40(20.20, 24.70)	23.00(20.50, 24.80)	0.433
LPA, mm	20.60(18.53, 22.68)	21.30(19.00, 23.20)	0.079
RVEDA, cm^2	23.50 ± 7.41	25.20 ± 8.76	**0.037**
RVESA, cm^2	14.80(10.60, 20.30)	15.50(11.00, 22.20)	0.376
FAC, %	34.45 ± 11.46	34.78 ± 11.05	0.783
TAPSE, mm	16.71 ± 4.00	16.60 ± 4.37	0.797
TAPSE/PASP	0.22(0.16, 0.39)	0.20(0.14, 0.38)	0.254
MPI	0.60(0.45, 0.80)	0.65(0.48, 0.85)	0.185
S', cm/s	10.70(9.00, 12.68)	11.00(8.38, 12.90)	0.909

Variables	Training cohort	Validation cohort	*p*-value
E', cm/s	7.00(5.50, 9.00)	8.00(6.42, 10.00)	**0.003**

Table 1
(continued)

Variables	Training cohort	Validation cohort	*p*-value
A', cm/s	12.10(9.03, 15.00)	13.50(9.68, 16.10)	**0.041**
RHC characteristics			
CVP, mmHg	5.00(3.00, 8.00)	5.00(3.00, 9.00)	0.359
sPAP, mmHg	69.73 ± 26.12	74.63 ± 26.92	0.076
dPAP, mmHg	26.00(16.00, 34.00)	27.00(20.00, 37.00)	0.274
PCWP, mmHg	8.28(6.00, 11.17)	9.00(7.00, 11.00)	0.362
CO, L/min	4.28(3.42, 5.21)	4.46(3.59, 5.50)	0.164
PVR, dyn·s·cm^{-5}	648.00(324.50, 974.00)	584.00(361.00, 1097.00)	0.788
CI, L/min	2.54(2.05, 3.00)	2.55(1.98, 2.89)	0.646

6MWD = six-minute walk distance; A' = late diastolic peak velocity of tricuspid annulus; BMI = body mass index; BSA = body surface area; CI = cardiac index; CO = cardiac output; CVP = central venous pressure; dBP = diastolic blood pressure; dPAP = diastolic pulmonary artery pressure; E' = early diastolic peak velocity of tricuspid annulus; EI = eccentricity Index; FAC = fractional area change; HR = heart rate; LPA = left pulmonary artery; LVEF = left ventricular ejection fraction; LVD = left ventricular diameter; MPI = myocardial performance index; MPA = main pulmonary artery; NT_proBNP = N-terminal pro-brain natriuretic peptide; PASP = pulmonary artery systolic pressure; PCWP = pulmonary capillary wedge pressure; PVR = pulmonary vascular resistance; RHC = right heart catheterization; RPA = right pulmonary artery; RVEDA = right ventricular end-diastolic area; RVESA = right ventricular end-systolic area; RVD = right ventricular diameter; S' = systolic peak velocity of tricuspid annulus; sBP = systolic blood pressure; sPAP = systolic pulmonary artery pressure; SpO2 = peripheral capillary oxygen saturation; TAPSE = tricuspid annular plane systolic excursion; WHO FC = world health organization functional class.

Based on the training cohort, eight ML models were constructed. Figure 3 presents the results of our ML models employing different classifiers. By comparing the performance of all classifiers on the validation cohort, the Naive Bayes model achieved the highest classification accuracy. The Naive Bayes model achieved an AUROC of 0.886, an accuracy of 0.736, a Brier score of 0.106, and an F1 score of 0.736, with an AUROC of 0.894 on the training cohort. Furthermore, its AUC values across different mPAP grading validation cohorts reached 0.994, 0.878, 0.779, and 0.892, respectively (Fig. 4A). Figures 4C-D present the corresponding confusion matrices for the test and validation cohorts' prediction results.

Table 2 presents the predictive performance of the Naive Bayes model across different mPAP severity grades. Sensitivity was highest for the normal group (92.9%), indicating the strongest ability to identify mPAP-normal samples; sensitivity was next highest for the severely elevated group (83.3%), indicating good detection of severe abnormalities; while sensitivity was moderate for the mildly elevated group (60.0%) and moderately elevated group (51.9%). However, specificity revealed consistently high levels across all groups (82.0%–98.1%), with the normal group exhibiting the highest specificity (98.1%). This demonstrates the model's strong ability to distinguish 'non-normal groups'. The normal group (86.7%) and the severely elevated group (82.0%) demonstrated higher accuracy, whereas the mildly elevated group (63.2%) and the moderately elevated group (53.8%) exhibited generally moderate accuracy.

Figure 5 displays the overall importance ranking of the six features in predicting mPAP classification, alongside their importance ranking within each classification. TAPSE/PASP emerges as the most significant predictor of mPAP, as evidenced by its elevated Shap value, underscoring its critical role in diagnosing pulmonary arterial hypertension. Moreover, PASP and EI were also significant predictors of mPAP classification. Within each group, the top two features for the normal group were EI and PASP; for the mildly elevated group, the top two were TAPSE/PASP and EI; while for the moderately and severely elevated groups, the most significant features were TAPSE/PASP and PASP. The strong similarity in feature importance across categories demonstrates the stability of these three features throughout the predictive process.

Table 2
Performance of Feature Fusion and Decision Fusion Models in Training Cohort and External Validation Cohort

Group	Model	mPAP ≤ 20 mmHg			20 mmHg<mPAP ≤ 35 mmHg			35 mmHg<mPAP ≤ 45 mmHg			mPAP>45 m
		Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision	Sensitivity
Training Cohort	LR	0.435	0.973	0.512	0.655	0.848	0.551	0.262	0.940	0.634	0.925
	RF	0.715	0.958	0.702	0.555	0.891	0.586	0.456	0.890	0.591	0.856
	GB	0.570	0.948	0.580	0.480	0.864	0.486	0.456	0.858	0.508	0.800
	SVM	0.800	0.933	0.609	0.516	0.898	0.586	0.500	0.844	0.511	0.769
	KNN	0.665	0.952	0.629	0.505	0.898	0.604	0.448	0.858	0.492	0.819
	NB	0.875	0.939	0.661	0.518	0.916	0.639	0.403	0.862	0.487	0.825
	XGB	0.570	0.955	0.575	0.505	0.875	0.524	0.511	0.851	0.523	0.788
	DT	0.720	0.948	0.675	0.470	0.847	0.433	0.458	0.815	0.470	0.681
Validation Cohort	LR	0.571	1.000	1.000	0.600	0.901	0.545	0.296	0.926	0.533	0.933
	RF	0.786	1.000	1.000	0.650	0.931	0.650	0.481	0.840	0.464	0.833
	GB	0.714	1.000	1.000	0.550	0.931	0.611	0.481	0.819	0.433	0.817
	SVM	0.929	0.981	0.867	0.550	0.941	0.647	0.630	0.819	0.500	0.783
	KNN	0.786	1.000	1.000	0.350	0.931	0.500	0.407	0.766	0.333	0.833
	NB	0.929	0.981	0.867	0.600	0.931	0.632	0.519	0.872	0.538	0.833
	XGB	0.643	1.000	1.000	0.600	0.911	0.571	0.444	0.830	0.429	0.817
	DT	0.571	0.991	0.889	0.550	0.901	0.524	0.519	0.798	0.424	0.783

DT = Decision Tree; GB = Gradient Boosting; KNN = K-Nearest Neighbors; LR = Logistic Regression; NB = Naive Bayes; RF = Random Forest; SVM = Support Vector Machine; XGB = Extreme Gradient Boosting.

Discussion

Precise grading prediction of mPAP holds significant importance for the clinical diagnosis and management of pulmonary arterial hypertension. In this study, we explored how ML methods, combined with clinical and TTE data, can be employed to achieve accurate mPAP grading. This research provides the first evidence that ML approaches can offer clinically actionable guidance for the precise grading of mPAP.

In this study, the Naive Bayes model demonstrated optimal performance, primarily due to its algorithmic properties grounded in Bayes' theorem and the assumption of feature conditional independence [27]. This model efficiently processes high-dimensional mixed features such as clinical and ultrasound characteristics, exhibits robust resilience to small sample sizes and noisy data, and effectively captures the nonlinear relationship between features and mPAP. From a clinical application perspective, the Naive Bayes model's high AUROC (0.886) and low Brier Score (0.106) indicate not only accurate classification but also reliable probabilistic predictions. This assists clinicians in assessing the confidence level of a patient's mPAP grade. Moreover, the model demonstrated excellent predictive performance in both the normal group and the group with severe elevation. Although sensitivity was moderate in the mildly elevated and moderately elevated groups (0.600 and 0.519, respectively), the specificity values of 0.820 and 0.867 demonstrate the model's strong ability to distinguish non-target categories. This effectively reduces the clinical risk of misclassifying normal individuals or patients with severe elevation as having mild/moderate elevation, providing a reliable reference for stratified disease diagnosis.

SHAP values are a widely recognized model interpretation method in ML, quantifying the weighting of each variable's influence on prediction outcomes [28]. SHAP analysis in this study revealed that TAPSE/PASP, PASP, and EI were the most significant predictors for mPAP grading. Among these, PASP—as an ultrasonically estimated pulmonary artery systolic pressure—often provides an intuitive reflection of pulmonary arterial pressure status, though it frequently carries inherent measurement error [29]. TAPSE reflects right ventricular longitudinal systolic function, while PASP directly represents pulmonary artery pressure. The ratio between these two quantifies the equilibrium between right ventricular functional reserve and pressure load [30]. This study found that TAPSE/PASP decreased significantly with increasing mPAP, consistent with the pathological mechanism [31]. In the early stages of PH, the right ventricle maintains normal TAPSE through compensatory systolic enhancement, with only a slight increase in PASP, hence a mild decrease in TAPSE/PASP. As the disease progresses, right ventricular myocardial remodeling leads to decompensated systolic function, resulting in decreased TAPSE, markedly elevated PASP, and a steep decline in TAPSE/PASP [32]. Consequently, TAPSE/PASP may serve as a combined functional-pressure marker for PH staging.

As a key predictor of mPAP grading, the core mechanism of EI lies in the elevation of mPAP, causing increased right ventricular pressure throughout the cardiac cycle [33]. The right ventricle exerts pressure on the interventricular septum during both systole and diastole, resulting in a ‘D’-shaped left ventricle throughout the cardiac cycle and progressively increasing EI [34]. As a pivotal morphological indicator, EI compensates for the limitations of pressure-based measures (PASP, TRV) and functional parameters in capturing morphological alterations. Its non-invasive acquisition via echocardiography and high

measurement reproducibility provide unique morphological evidence for mPAP stratification, significantly enhancing the model's discriminatory power across different grades—particularly within the normal group.

However, this study also has certain limitations. Firstly, as a single-center study lacking external validation, it may exhibit selection bias. Multicentre, large-sample studies are required to validate the model's generalizability. Secondly, the study predominantly included patients with PH classified as Type I and Type IV, with fewer participants from other subtypes. Consequently, the model's applicability across the entire spectrum of PH subtypes may be limited, potentially failing to meet clinical demands for screening and grading patients with diverse etiologies. Thirdly, the model incorporates primarily conventional clinical and TTE indicators, excluding other imaging examinations and serological markers as potential predictors. This may omit key pathophysiological information relevant to mPAP grading, limiting further enhancement of the model's predictive efficacy. Future research should expand sample sizes to include patients across all PH subtypes, conduct multicenter prospective validation studies, and integrate multidimensional biomarkers with radiomic features to optimize the model's generalizability and accuracy.

Conclusions

The ML model developed in this study is based on clinical and echocardiographic data and can effectively predict mPAP grading. This non-invasive predictive model is a reliable tool for clinical PH staging, offering the potential to reduce reliance on invasive RHC procedures and support precise therapeutic decision-making.

Abbreviations

6MWD
6-minute walk distance
AUC
area under the receiver operating characteristic curve
BSA
body surface area
CO
cardiac output
CVP
central venous pressure
dPAP
diastolic pulmonary artery pressure
DT
decision tree
EI
eccentricity index
FAC
fractional area change
GB
gradient boosting
HR
heart rate
KNN
K-nearest neighbors
Lasso
least absolute shrinkage and selection operator
LPA
left pulmonary artery diameter
LVD
left ventricular diameter
LVEF
left ventricular ejection fraction
LR
logistic regression
mPAP
mean pulmonary arterial pressure
ML
machine learning
MPI
myocardial performance index
MPA

main pulmonary artery diameter
NB
naive Bayes
NT-proBNP
N-terminal pro-B-type natriuretic peptide
PCWP
pulmonary capillary wedge pressure
PH
pulmonary hypertension
PASP
systolic pulmonary artery pressure
PVR
pulmonary vascular resistance
RHC
right heart catheterization
RPA
right pulmonary artery diameter
RVD
right ventricular diameter
RVEDA
right ventricular end-diastolic area
RVESA
right ventricular end-systolic area
RF
random forest
SHAP
SHapley additive exPlanations
SMOTE
synthetic minority over-sampling technique
SpO_2
oxygen saturation
sPAP
systolic pulmonary artery pressure
SVM
support vector machine
TAPSE
tricuspid annular plane systolic excursion
TTE
transthoracic echocardiography
XGB
extreme gradient boosting

Declarations

Ethics approval and consent to participate

Ethics approval and consent to participate This retrospective study was approved by the institutional ethics committee of Beijing Chaoyang hospital, and the requirement for informed consent was waived due to the use of anonymized patient data. The study was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

supported by Beijing Research Ward Excellence Program, No.BRWEP2024W112030103;the National Natural Science Foundation of China, No.82572244Beijing Municipal Administration of Hospitals Incubating Program, No. Z221100007422028.

Author Contribution

1. Conception and design: XPD, YDL. (II) Administrative support: YDL, XPD. (III) Provision of study materials or patients: JNG, SQY, YHY,QMX. (IV) Collection and assembly of data: XPD. (V) Data analysis and interpretation: XPD, YDL, JYH, RF, DCG, XYZ. (VI) Manuscript writing: All authors. (VII) Final approval of manuscript: All authors.

Acknowledgement

We thank all the patients for their participation and the clinical and research teams for their contributions to data collection and study coordination.

Data Availability

All data used and analyzed during this study are included in the manuscript.

References

1. Humbert M, Kovacs G, Hoeper MM, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Heart J. 2022;43(38):3618–731. 10.1093/eurheartj/ehac237.
2. Kovacs G, Bartolome S, Denton CP, et al. Definition, classification and diagnosis of pulmonary hypertension. Eur Respir J. 2024;64(4):2401324. 10.1183/13993003.01324-2024.
3. Mukherjee M, Rudski LG, Addetia K, et al. Guidelines for the Echocardiographic Assessment of the Right Heart in Adults and Special Considerations in Pulmonary Hypertension: Recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr Off Publ Am Soc Echocardiogr. 2025;38(3):141–86. 10.1016/j.echo.2025.01.006.
4. Mocumbi A, Humbert M, Saxena A, et al. Pulmonary hypertension. Nat Rev Dis Primer. 2024;10(1):1. 10.1038/s41572-023-00486-7.
5. Gašparović K, Pavliša G, Hrabak Paar M, et al. Diagnostic accuracy, sensitivity, and specificity of CT pulmonary artery to aorta diameter ratio in screening for pulmonary hypertension in end-stage COPD patients. Croat Med J. 2021;62(5):446–445. 10.3325/cmj.2021.62.446.
6. Mukherjee M, Rudski LG, Addetia K, et al. Guidelines for the Echocardiographic Assessment of the Right Heart in Adults and Special Considerations in Pulmonary Hypertension: Recommendations from the American Society of Echocardiography. J Am Soc Echocardiogr Off Publ Am Soc Echocardiogr. 2025;38(3):141–86. 10.1016/j.echo.2025.01.006.
7. Ley L, Grimminger F, Richter M, Tello K, Ghofrani A, Bandorski D. The Early Detection of Pulmonary Hypertension. Dtsch Arzteblatt Int. 2023;120(48):823–30. 10.3238/arztebl.m2023.0222.
8. Hl A, Rk MS. Pulmonary Hypertension in Interstitial Lung Disease: A Systematic Review and Meta-Analysis. Chest. 2024;166(4). 10.1016/j.chest.2024.04.025.
9. Yang S, Lei S, Peng F, Wu S-J. Detection of Pulmonary Hypertension by Combining Echocardiography and Chest Radiography. Acad Radiol. 2022;29(Suppl 2):S23–30. 10.1016/j.acra.2020.10.003.
10. Jiang T, Gradus JL, Rosellini AJ. Supervised Machine Learning: A Brief Primer. Behav Ther. 2020;51(5):675–87. 10.1016/j.beth.2020.05.002.
11. Kovács A, Tokodi M. Refining Echocardiographic Surveillance of Aortic Stenosis Using Machine Learning: Toward Personalized and Sustainable Follow-Up Schemes. JACC Cardiovasc Imaging. 2023;16(6):745–8. 10.1016/j.jcmg.2023.01.019.
12. Upton R, Mumith A, Beqiri A, et al. Automated Echocardiographic Detection of Severe Coronary Artery Disease Using Artificial Intelligence. JACC Cardiovasc Imaging. 2022;15(5):715–27. 10.1016/j.jcmg.2021.10.013.
13. Diaz DJ, Kulikova AV, Ellington AD, Wilke CO. Using machine learning to predict the effects and consequences of mutations in proteins. Curr Opin Struct Biol. 2023;78:102518. 10.1016/j.sbi.2022.102518.
14. Shelley B, Shaw M. Machine learning and preoperative risk prediction: the machines are coming. Br J Anaesth. 2024;133(5):925–30. 10.1016/j.bja.2024.07.015.
15. Sammut S-J, Crispin-Ortuzar M, Chin S-F, et al. Multi-omic machine learning predictor of breast cancer therapy response. Nature. 2022;601(7894):623–9. 10.1038/s41586-021-04278-5.
16. Zhu L, Pan J, Mou W, et al. Harnessing artificial intelligence for prostate cancer management. Cell Rep Med. 2024;5(4):101506. 10.1016/j.xcrm.2024.101506.
17. Guo Y, Xia C, Zhong Y, et al. Machine learning-enhanced echocardiography for screening coronary artery disease. Biomed Eng Online. 2023;22(1):44. 10.1186/s12938-023-01106-x.
18. Chao C-J, Kato N, Scott CG, et al. Unsupervised Machine Learning for Assessment of Left Ventricular Diastolic Function and Risk Stratification. J Am Soc Echocardiogr Off Publ Am Soc Echocardiogr. 2022;35(12):1214–e12258. 10.1016/j.echo.2022.06.013.

19. Xi Q, Gong J, Wang J, et al. Machine learning in CTEPH: predicting the efficacy of BPA based on clinical and echocardiographic features. BMC Med Imaging. 2025;25(1):328. 10.1186/s12880-025-01870-3.
20. Hirata Y, Tsuji T, Kotoku J, Sata M, Kusunose K. Echocardiographic artificial intelligence for pulmonary hypertension classification. Heart Br Card Soc. 2024;110(8):586–93. 10.1136/heartjnl-2023-323320.
21. Zhao W, Huang Z, Diao X, et al. Development and validation of multimodal deep learning algorithms for detecting pulmonary hypertension. NPJ Digit Med. 2025;8(1):198. 10.1038/s41746-025-01593-3.
22. Anand V, Weston AD, Scott CG, Kane GC, Pellikka PA, Carter RE. Machine Learning for Diagnosis of Pulmonary Hypertension by Echocardiography. Mayo Clin Proc. 2024;99(2):260–270. 10.1016/j.mayocp.2023.05.006
23. Kogan E, Didden E-M, Lee E, et al. A machine learning approach to identifying patients with pulmonary hypertension using real-world electronic health records. Int J Cardiol. 2023;374:95–9. 10.1016/j.ijcard.2022.12.016.
24. Celestin B, Bagherzadeh SP, Santana E et al. Artificial Intelligence-Based Echocardiography in Pulmonary Arterial Hypertension. Chest. 2025;S0012-3692(25)05122-0. 10.1016/j.chest.2025.06.052
25. Qin D, Yang S, Lin L, et al. Clinical research on echocardiographic screening for pulmonary hypertension: TRV > 2.8 m/s and mPAP > 20 mmHg. Quant Imaging Med Surg. 2025;15(10):9792–804. 10.21037/qims-2024-2878.
26. Sun T, Liu J, Yuan H, Li X, Yan H. Construction of a risk prediction model for lung infection after chemotherapy in lung cancer patients based on the machine learning algorithm. Front Oncol. 2024;14:1403392. 10.3389/fonc.2024.1403392.
27. Duan HN, Hearne G, Polikar R, Rosen GL. The Naïve Bayes classifier + + for metagenomic taxonomic classification-query evaluation. Bioinforma Oxf Engl. 2024;41(1):btae743. 10.1093/bioinformatics/btae743.
28. Song Y, Zhang D, Wang Q, et al. Prediction models for postoperative delirium in elderly patients with machine-learning algorithms and SHapley Additive exPlanations. Transl Psychiatry. 2024;14(1):57. 10.1038/s41398-024-02762-w.
29. Fortmeier V, Lachmann M, Körber MI, et al. Solving the Pulmonary Hypertension Paradox in Patients With Severe Tricuspid Regurgitation by Employing Artificial Intelligence. JACC Cardiovasc Interv. 2022;15(4):381–94. 10.1016/j.jcin.2021.12.043.
30. Tello K, Wan J, Dalmer A, et al. Validation of the Tricuspid Annular Plane Systolic Excursion/Systolic Pulmonary Artery Pressure Ratio for the Assessment of Right Ventricular-Arterial Coupling in Severe Pulmonary Hypertension. Circ Cardiovasc Imaging. 2019;12(9):e009047. 10.1161/CIRCIMAGING.119.009047.
31. Conde-Camacho R, Tuta-Quintero E, Varón-Vega F, et al. Association of hemodynamic and functional variables with pulmonary vasculopathy in lung transplant recipients living at high altitude: A retrospective study. Sci Prog. 2025;108(3):368504251367283. 10.1177/00368504251367283.
32. Rako ZA, Kremer N, Yogeswaran A, Richter MJ, Tello K. Adaptive versus maladaptive right ventricular remodelling. ESC Heart Fail. 2023;10(2):762–75. 10.1002/ehf2.14233.
33. Butt MU, Jabri A, Hamade H, et al. Predicting the Severity and Outcome of Persistent Pulmonary Hypertension of the Newborn Using New Echocardiography Parameters. Curr Probl Cardiol. 2023;48(8):101181. 10.1016/j.cpcardiol.2022.101181.
34. Madonna R, Tocci G, Biondi F, Cipollini V, Morganti R, De Caterina R. Chronic Thromboembolic Pulmonary Disease: Right Ventricular Function and Pulmonary Hemodynamics in a 4-Year Follow-Up. Int J Mol Sci. 2025;26(21):10617. 10.3390/ijms262110617.

Figures

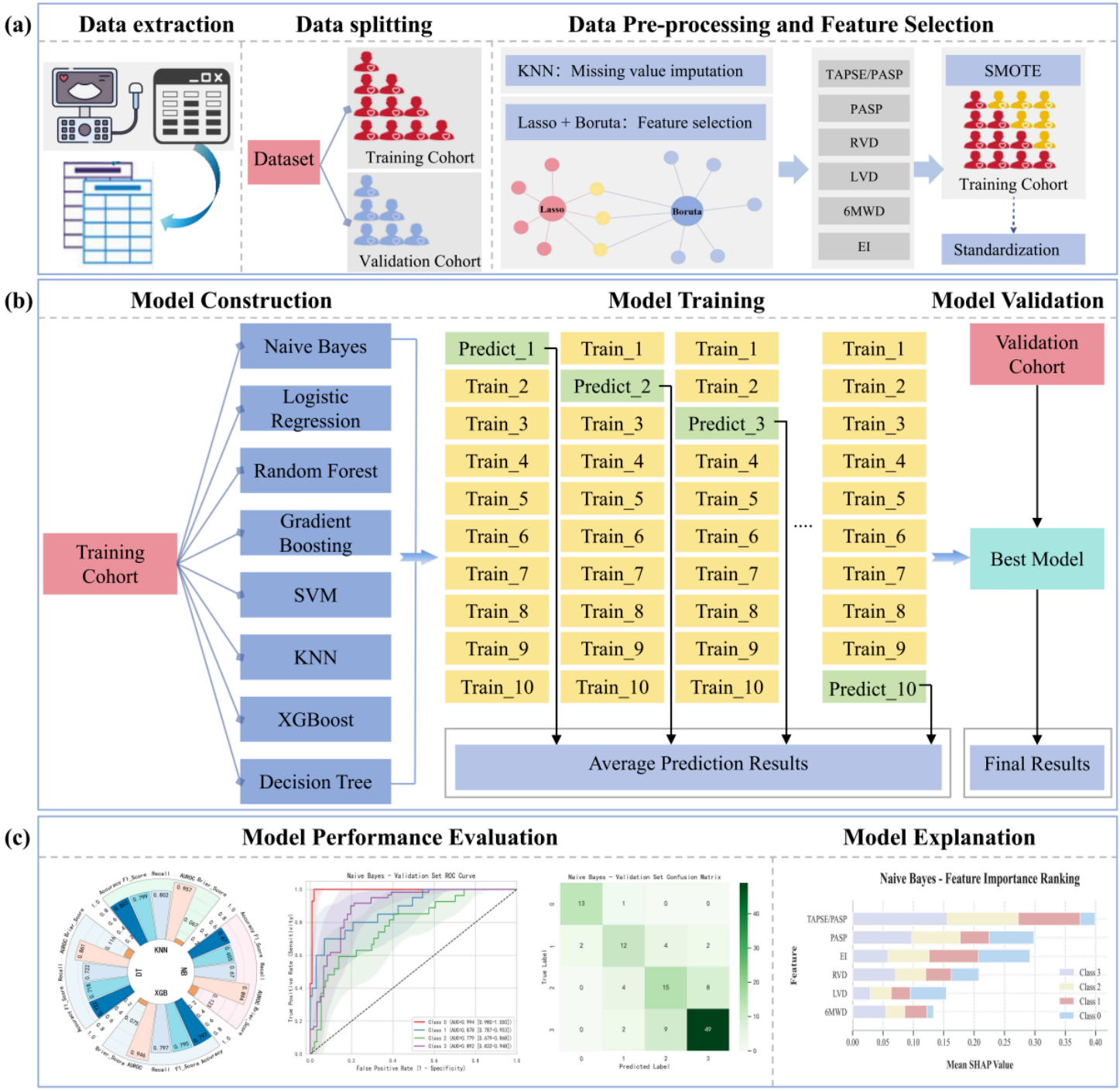


Figure 1

The overall process of this study. (a) Data Extraction, preprocessing and feature selection. (b) Model Construction and Validation. (c) Model performance evaluation and interpretation. 6MWD = six-minute walk distance; EI = eccentricity Index; KNN = K-Nearest Neighbors; LVD = left ventricular diameter; PASP = pulmonary artery systolic pressure; RVD = right ventricular diameter; SMOTE = synthetic minority over-sampling technique; SVM = Support Vector Machine; TAPSE = tricuspid annular plane systolic excursion.

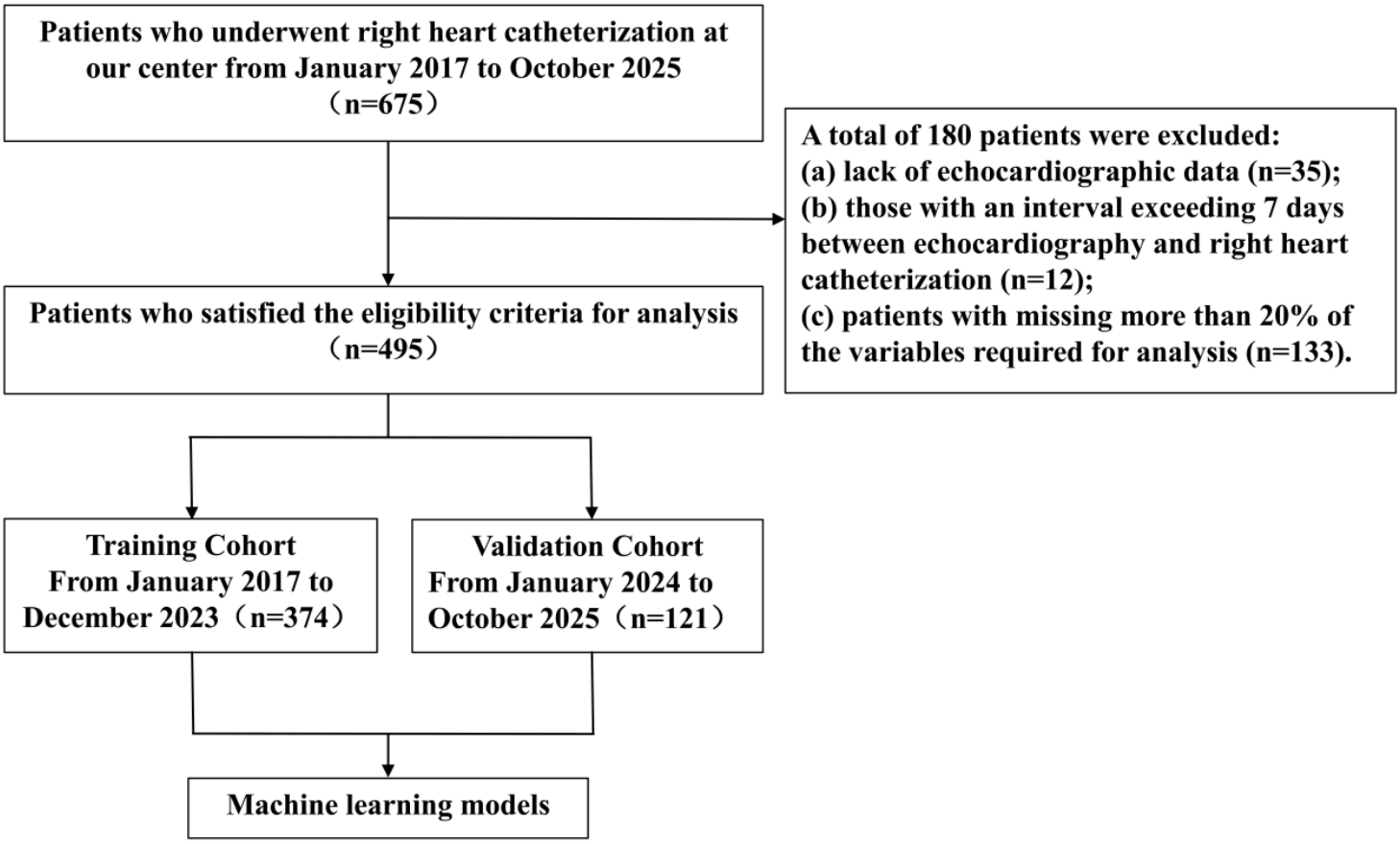


Figure 2

Patient selection flowchart

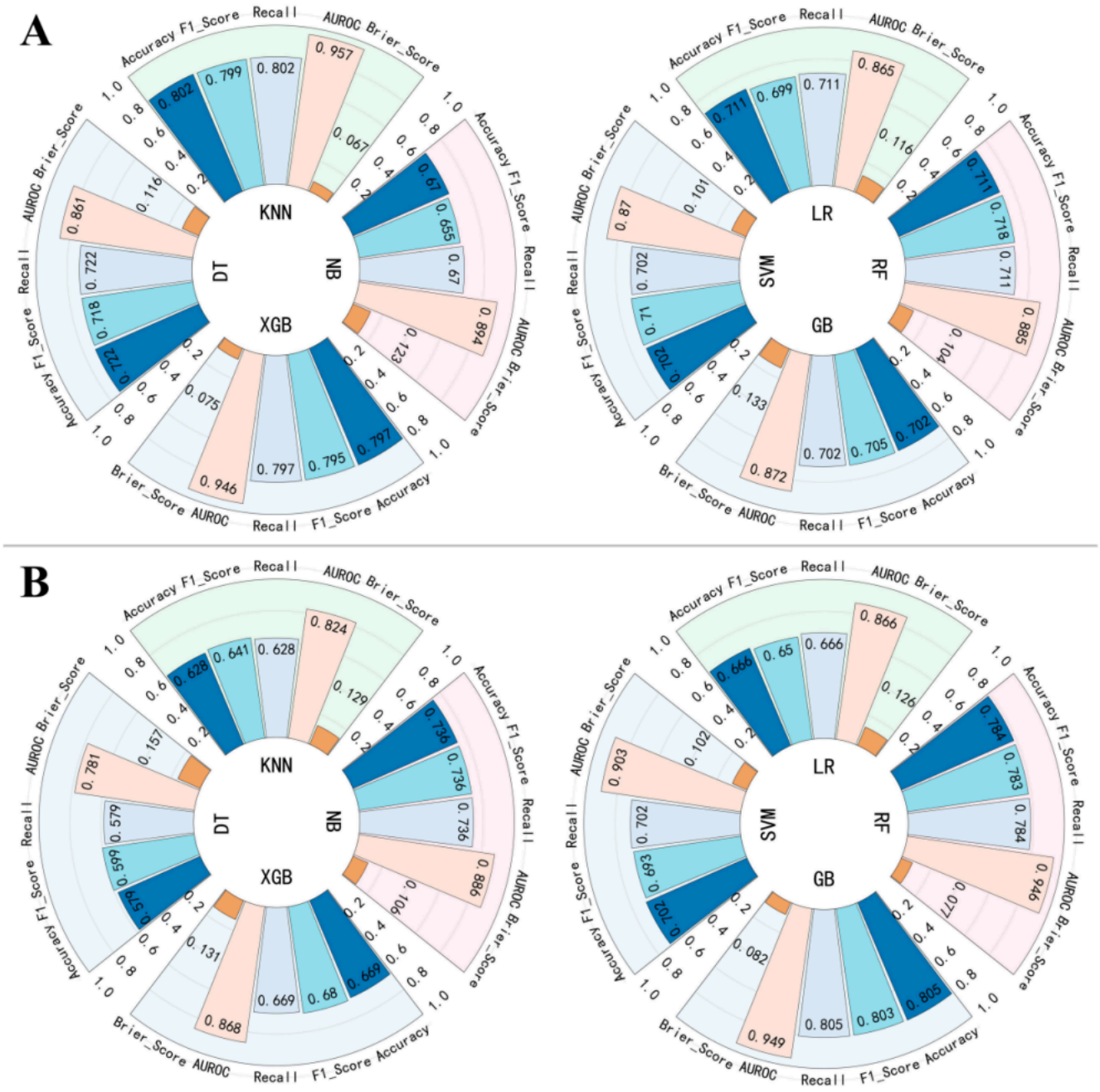


Figure 3

The performance of all models in the training cohort (A) and the validation cohort (B). DT = Decision Tree; GB = Gradient Boosting; KNN = K-Nearest Neighbors; LR = Logistic Regression; NB = Naive Bayes; RF = Random Forest; SVM = Support Vector Machine; XGB = Extreme Gradient Boosting.

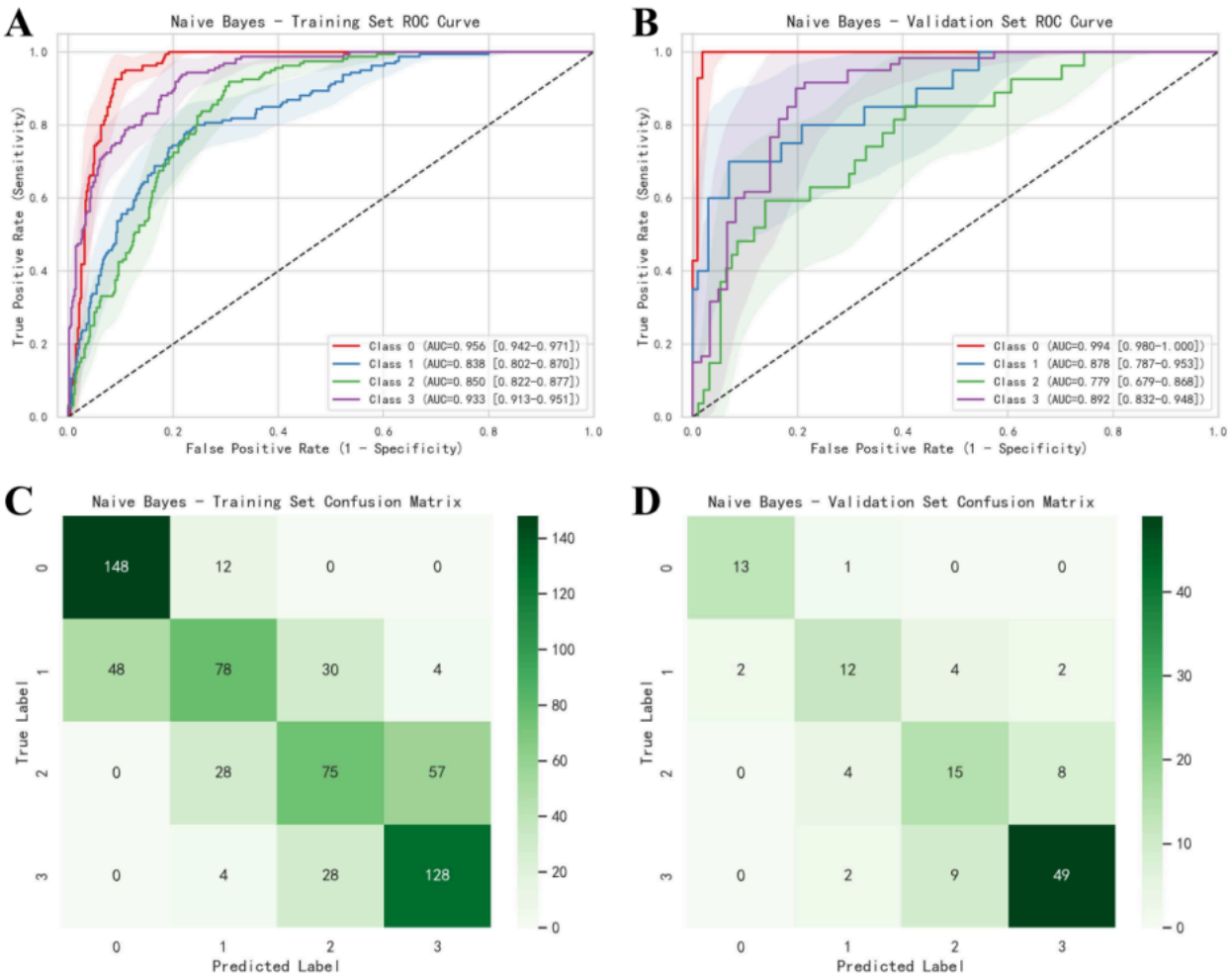


Figure 4

Receiver operating characteristic (ROC) curves of the Naive Bayes model in the training (A) and validation (B) cohorts; confusion matrices of the model in the training (C) and validation (D) cohorts. Class 0 = mPAP ≤ 20 mmHg; Class 1 = 20 mmHg < mPAP ≤ 35 mmHg; Class 2 = 35 mmHg < mPAP ≤ 45 mmHg; Class 3 = mPAP > 45 mmHg.

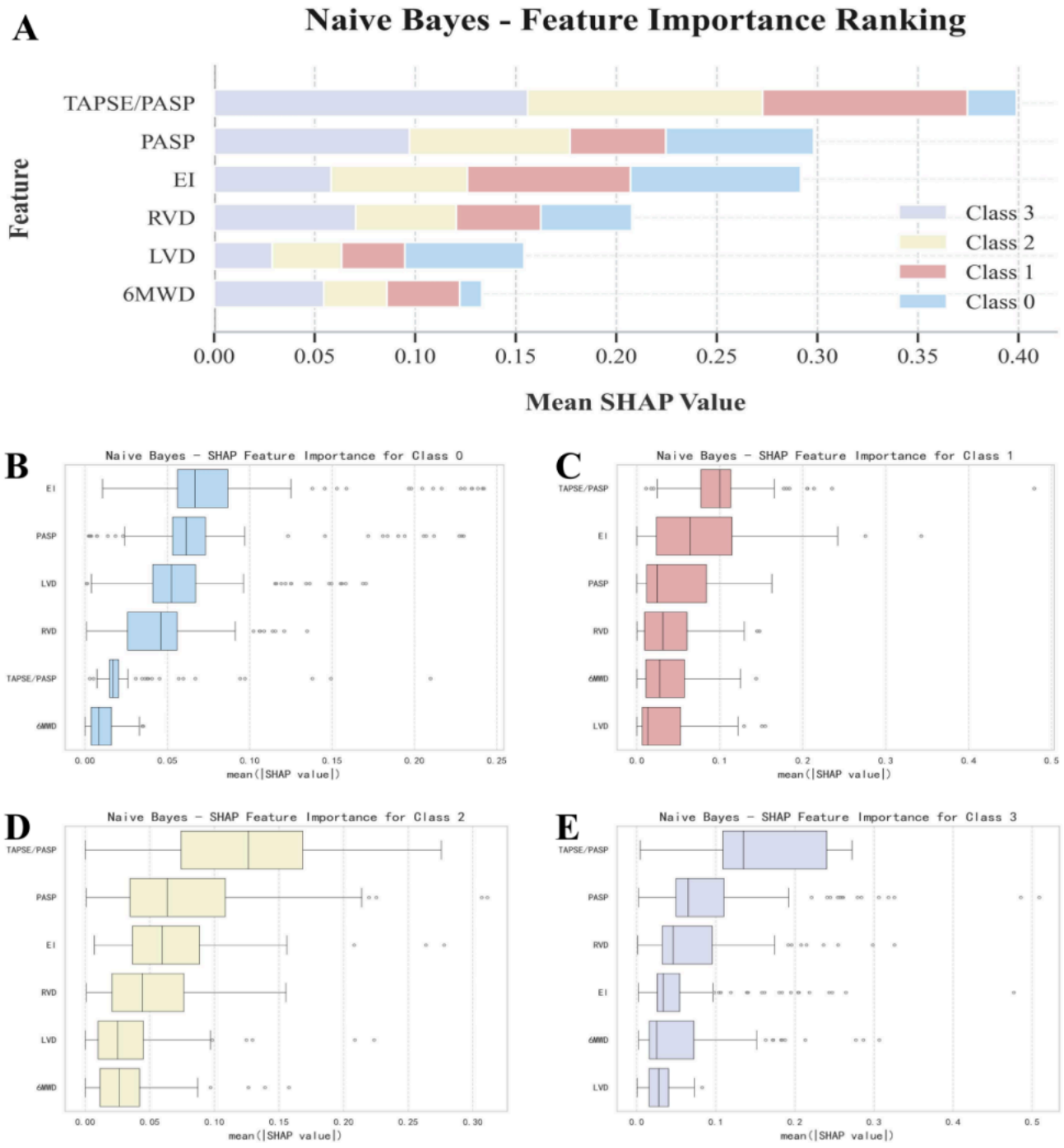


Figure 5

SHAP interpretation of the Naive Bayes model. (A) Ranking of predictor variables by SHAP importance scores. (B-E) Ranking of predictor variables by SHAP importance scores across each category. Class 0 = mPAP ≤20 mmHg; Class 1 = 20 mmHg < mPAP ≤35 mmHg; Class 2 = 35 mmHg < mPAP ≤45 mmHg; Class 3 = mPAP > 45 mmHg; SHAP=SHapley Additive exPlanations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.docx