Original Article | J Adv Med Biomed Res. 2023; 31(146): 210-220

Journal of Advances in Medical and Biomedical Research | ISSN:2676-6264

Assessment of Machine Learning Approaches to Predict in-Hospital Mortality in Patients Underwent Prosthetic Heart Valve Replacement Surgery

Maryam Shojaeifard[1], Hassan Ahangar[2*], Sepehr Gohari[3,4], Mehrdad Oveisi[5], Majid Maleki[1], Tara Reshadmanesh[3], Shahram Arsang-Jang[6], Mahsa Mahjani[7], Mozhgan Pourkeshavarz[8], Ghasem Hajianfar[9], Saeedeh Mazloomzadeh[8], Isaac Shiri[9], Sheida Gohari[10]

1. *Echocardiography Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran*
2. *Dept. of Cardiology, Mousavi Hospital, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran*
3. *Student Research Center, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran*
4. *Dept. of Family Medicine, Alborz University of Medical Sciences, Alborz, Iran*
5. *Comprehensive Cancer Centre, School of Cancer & Pharmaceutical Sciences, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom*
6. *Dept. of Biostatistics, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran*
7. *School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
8. *Dept. of Biomedical and Health Informatics, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran*
9. *Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland*
10. *Dept. of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, New York, United States*

Article Info

10.30699/jambs.31.146.210

Received: 2022/12/05;
Accepted: 2023/04/09;
Published Online: 26 June 2023;

Use your device to scan and read the article online

Corresponding Information:
Hassan Ahangar,
Dept. of Cardiology, Mousavi Hospital, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

E-Mail: hassan.ahangar@yahoo.com

ABSTRACT

Background & Objective: Machine learning and artificial intelligence are useful tools to analyze data with multiple variables. It has been shown that the prediction models obtained by Machine learning have better performance than the conventional statistical methods. This study was aimed to assess the risk factors and determine the best machine learning prediction model/s for in-hospital mortality among patients who underwent prosthetic valve replacement surgery.

Materials & Methods: In this retrospective cross-sectional study, patient's pre-operative, intra-operative and post-operative data underwent univariate analysis. Feature importance determination was carried out using algorithms including principal component analysis (PCA), support vector machine (SVM), random forest (RF) model-based, and recursive feature elimination (RFE). Then, 13 machine learning classifiers were implemented for in-hospital prediction model.

Results: The In-hospital mortality rate was 6.36%. Data from 2455 patients underwent final analysis. The machine learning results revealed that among pre-operative features, Adaptive boost (AB) and RF classifiers (AUC: 0.82±0.033; 0.78±0.028, respectively); among intra-operative features, AB and K-nearest neighbors (KNN) classifiers (AUC: 0.68±0.014); among postoperative features, AB and RF classifiers (AUC: 0.9±0.1; 0.88±0.095, respectively); and among all features, AB and LR classifiers (AUC: 0.93±0.049; 0.93±0.055, respectively) had the best performance in prediction of in-hospital mortality.

Conclusion: The AB classifier was determined as the best model in prediction of in-hospital mortality in all 4 datasets.

Keywords: Prosthetic valve replacement, In-hospital mortality, Risk factor, Machine learning

Introduction

Valvular heart disease (VHD) is one of the most prevalent cardiac diseases affecting about 2.5% of the population in the United States (1).

The survival advantage of heart valve replacement makes this method the treatment of choice for patients with severe valve stenosis, regurgitation, or high-degree valve calcification who develop symptoms (2). Such invasive procedures have several complications, such as various comorbidities and higher risk of mortality; therefore, proper risk assessment must be carried out. On the other hand, outcomes and mortality rates vary depending on surgical techniques, valve location, and patient characteristics (3,4).Previous studies have investigated independent risk factors in association with mortality after valve replacement, but they have included a limited number of cases due to fewer surgical treatments for valve replacement than other procedures, e.g., coronary artery bypass graft (CABG) surgery (5,6).

Artificial intelligence and machine learning approaches are new methods consisting of a collection of technologies for analyzing huge amounts of data, which allow the health care system to find algorithms and models for diagnosing diseases and predicting risks and complications during various medical procedures (7). These novel methods of analysis may reveal new prediction models (8). Identifying more robust prediction models helps cardiac surgeons and hospital staff to choose the best management approach and monitor patients with a higher risk of mortality more closely.

In the present study,we collected a significant amount of clinical data related to patients undergoing heart valve replacement surgery at a large referral tertiary hospital center and assessed a variety of machine learning models in order to determine which models are the most effective at predicting in-hospital mortality in patients who underwent prosthetic valve replacement surgery.

Materials and Methods

Study design and setting

This study was a single center retrospective cross-sectional assessment of patients who underwent prosthetic valve replacement surgery (pooled population of single and multiple valve replacement) from March 2009 to March 2017 in Rajaei Heart Centre, Tehran, Iran. The study was conducted in accordance with the principles of the declaration of Helsinki and approved by the Ethic Committee of Iran University of medical sciences (June 7 2020; ID: IR.RHC.REC.1399.020). The informed consent was waived due to the retrospective design of the study.

Variables and outcomes

We retrospectively reviewed and collected the electronic data of the study population. The source of the data which was retrieved from the patients' medical records at the pre/intra and post operation stages. Variables were comprised of patient demographics, laboratory findings, and surgery associated variables. Patients with extensive amount of missing data were excluded from the study. The primary outcome was the incidence of in-hospital mortality. The secondary outcome was to investigate the best prediction model of in-hospital mortality using machine learning method.

Statistical analysis

Risk factor assessments

We used Spearman correlation matrix to assess the correlation between study variables and in-hospital mortality. The variance inflation factor (VIF) was used to determine possible collinearity among independent variables, values >2 considered collinearity problem. Variables with significant association in univariate analysis included to final regression model (Multivariate analysis) to control the confounding effects. We used Multiple Bayesian logistic regression model to examine the association between study variables and in-hospital mortality. Risk of mortality was measured using a Bayesian Odds Ratio (OR= Exp [Beta]) and a 95% credible interval (CI).

The Receiver operating characteristic was conducted to predictors with statistically significant association in multiple regression model to determine optimal classification of mortality.

Suture type, international normalized ratio (INR) level After 4 days of warfarin initiation, last blood urea nitrogen (BUN), Partial thromboplastin time (PTT), Hemoglobin (Hb), Platelet (PLT) Count, Creatinine (Cr), Sodium (Na), Potassium (K), Proteinuria, and valve replacement site were included in the model. The term "last" or "end" in this study was defined as the last laboratory result before discharge or death. The Bernoulli prior distribution was used in the Multiple Bayesian logistic regression model. The Rstan, brms, and ggplot2 packages were used to implement analysis in R 4.03 software. A P-value<0.05 was considered statistically significant (9, 10).

Machine learning considerations

Altogether, data of 2455 patients underwent considered for the assessments. Different collected features were divided into the three main category of: pre-operative, intra-operative and post-operative features. Four sub-datasets were used to predict mortality including pre-operative, intra-operative and post-operative, and all features.

Features Importance

Feature importance was carried out using four algorithms including: Principal component analysis (PCA), support vector machine (SVM) model-based, random forest (RF) model-based, and recursive feature elimination (RFE). Each method gives a score for features and the sum score of the four methods was used in this study. The final sum score was used for model evaluation.

Machine Learning Classifiers

In this study, 13 classifiers were implemented for in-hospital mortality prediction.These methods included ensemble learning methods [adaptive boosting (AB), bagging (BAG), and random forest (RF)], naïve Bayes (NB) models [Bernoulli naïve Bayes (BNB), Gaussian naïve Bayes (GNB) and Multinomial naïve Bayes (MNB)], generalized linear models [logistic regression (LR) and stochastic gradient descent (SGD)], support vector machine (SVM), nearest neighbor's model (k-nearest neighbors, KNN), quadratic discriminant analysis model (QDA), decision trees (DT) model (C5.0), and multi-layer perception (MLP) model. Details of each model's hyper-parameter are shown in Table 1.

Table 1. Machine learning methods applied in this study.

Machine learning methods	Machine learning algorithm	Abbreviation
Ensemble learning	Adaptive boosting	AB
	Bagging	BAG
	Random forest	RF
naïve Bayes models	Gaussian naïve Bayes	GNB
	Multinomial naïve Bayes	MNB
Generalized linear models	Logistic regression	LR
	Stochastic gradient descent	SGD
Support vector machine model	Support vector machine	SVM
Nearest neighbor's model	K-nearest neighbors	KNN
Quadratic discriminant analysis model	Quadratic discriminant analysis	QDA
Decision trees model	Decision trees (C5.0)	DT
Multi-layer perception model	Multi-layer perception	MLP

Model Design

Based on sorted feature importance, we prepared different dataset with different feature number (from one feature to all features). Train classifiers performed with ten cross-validation (10 cross-validation sample is randomly partitioned into ten equal size subsamples. Of the 10 subsamples, a single subsample is retained as the test data, while the remaining nine subsamples are used as training data, repeated for 10 times). For each dataset 13 classifiers were trained. In pre-operative dataset we had 40 features and also prepared 40 datasets with train and test. The cross combination of number of features and classifiers for this dataset resulted 520 models. Intra-operative dataset had 11 features and 143 models. Post-operative dataset had 21 features and 273 models. The last dataset with all features had 72 features and 936 models were obtained. All data analysis include feature selection and classifier was performed in in-house developed python framework in open-source python library Scikit-Learn (11).

Multivariate analysis

All data were divided between training (1963 patient) and test (491 patients) sets, and all evaluation performed on unseen test datasets (491 patient). Performance of models in test dataset were evaluated via computing the receiver operator characteristic (ROC), the area under the curve (AUC), sensitivity (SEN), specificity (SPE), positive predictive values (PPV) and negative predictive values (NPV).

Results

Through the assessments, 5076 patients underwent prosthetic valve replacement. Among them 323 (6.36%) patients were died before discharge. After excluding the patients with extensive missing data, data of the 2455 patients were included into the prediction models.

Univariate Analysis

The results of univariate analysis are shown in Table 2.

Table 2. The results of predictors of in-hospital mortality: The univariate analysis

Variable	OR	SE	P-value	95% CI for OR
Age (year)	1.048	0.004	<0.001	[1.040, 1.060]
Gender (Female/Male)	0.895	0.110	0.359	[0.710, 1.130]
Replced valve site (pulmonary)*	-	-	-	-
Aorta, Mitral or both	3.887	1.000	<0.001	[2.350, 6.420]
Tricuspid	7.159	2.230	<0.001	[3.890, 13.170]
Suture type	-	-	-	-

Variable	OR	SE	P-value	95% CI for OR
satureless/individual	13.119	2.300	<0.001	[9.300, 18.510]
satureless/continuose	2.028	0.290	<0.001	[1.530, 2.690]
Decortication (Yes/No)	0.982	0.120	0.880	[0.770, 1.240]
CPB/Pump time (min)	1.013	0.001	<0.001	[1.010, 1.020]
K_end	12.002	1.590	<0.001	[9.260, 15.550]
K_before	1.456	0.160	0.001	[1.170, 1.820]
Cr_before	2.772	0.360	<0.001	[2.140, 3.590]
Cr_end	6.826	0.790	<0.001	[5.440, 8.560]
Mg_end	1.652	0.210	<0.001	[1.280, 2.120]
UA	1.317	0.040	<0.001	[1.240, 1.390]
INR After 4 days	1.229	0.060	<0.001	[1.130, 1.340]
INR_end	1.015	0.041	0.705	[0.940, 1.100]
Na_end	1.212	0.010	<0.001	[1.180, 1.240]
BUN_before	1.056	0.010	<0.001	[1.050, 1.070]
BUN_end	1.103	0.010	<0.001	[1.090, 1.110]
FBS_before	1.007	0.001	<0.001	[1.004, 1.010]
PTT	1.043	0.003	<0.001	[1.040, 1.050]
Na_before	0.919	0.010	<0.001	[0.893, 0.946]
ESR_before	1.024	0.003	<0.001	[1.020, 1.030]
ESR_end	0.983	0.003	<0.001	[0.978, 0.988]
LDH	1.001	0.000	<0.001	[1.001, 1.002]
Cholesterol	0.991	0.002	<0.001	[0.988, 0.994]
LDL	0.989	0.002	<0.001	[0.985, 0.993]
Triglyceride	0.998	0.001	0.027	[0.996, 1.000]
HDL	0.975	0.010	<0.001	[0.962, 0.988]
Plt Count_before	0.995	0.001	<0.001	[0.993, 0.997]
Plt Count_end	0.969	0.002	<0.001	[0.966, 0.972]
Hb_before	0.762	0.020	<0.001	[0.702, 0.810]
Hb_end	0.347	0.020	<0.001	[0.310, 0.390]
Hct_before	0.940	0.010	<0.001	[0.919, 0.961]
Hct_end	0.774	0.010	<0.001	[0.750, 0.800]
Ca_end	0.561	0.050	<0.001	[0.480, 0.660]
Prosthetic valve type	1.572	0.250	0.004	[1.150, 2.140]
CABG (yes/no)	3.060	0.440	<0.001	[2.310, 4.060]
Protein				
1+	2.082	0.410	<0.001	[1.420, 3.060]
2+	8.480	2.030	<0.001	[5.300, 13.560]

CPB time: Cardio-pulmonary bypass time; K: Potassium; Cr: Creatinine; Mg: Magnesium; UA: Uric Acid; INR: International normalized ratio; Na: Sodium; BUN: Blood urea nitrogen; FBS: Fasting blood sugar, PTT: Partial thromboplastin time; Na: Sodium; ESR: Erythrocyte sedimentation rate; LDH: Lactate dehydrogenase; LDL: Low-density lipoprotein cholesterol; HDL: High-density lipoprotein cholesterol; Plt: Platelet; Hb: Hemoglobin; Hct: Hematocrit; Ca: Calcium; CABG: Coronary artery bypass surgery.

Multivariate analysis for the risk factor assessment

Correlations among variables and possible collinearity are presented in Figure 1. We excluded the variables with high collinearity from regression mode.

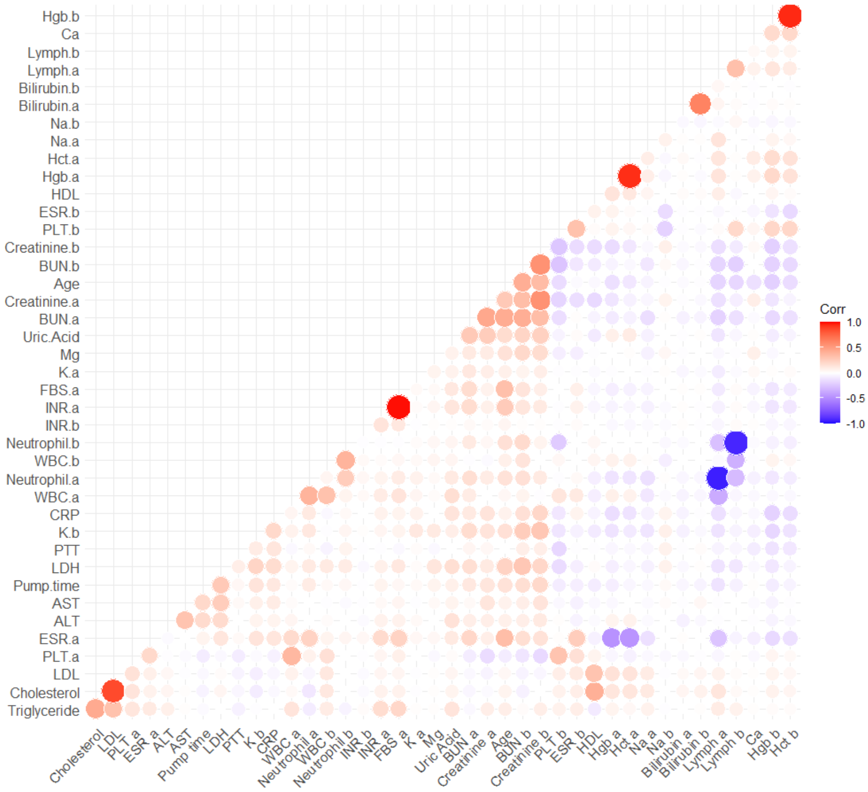

Figure 1. Correlations among variables. ALT: Alanine transaminase; AST: Aspartate transferase; CRP: C-reactive protein, the rest of abbreviations are similar to Table 2.

According to the results of final multiple Bayesian logistic regression model, the positive statistically significant association was observed between in-hospital mortality rate and Cardiopulmonary bypass (Pump, CPB) time per minute (OR=1.01), serum BUN (OR=1.102), PTT (OR=1.025), and serum Na (OR=1.173) variables. While, the inverse statistically significant association was observed between mortality rate and Hemoglobin (OR=0.606), PLT (OR=0.973), and serum creatinine (OR=0.618) variables.

Compared to negative proteinuria, individuals with proteinuria 2+ have 9.92 times higher risk of mortality, while we did not find this association in proteinuria 1+. The serum potassium significantly increased in-hospital mortality rate by 7.91 (95% CI: 4.43, 15.3). A positive association was observed between in-hospital mortality and valve replacement site (OR=2.12; 95% CI: 3.22, 1088.6). In the regression analysis, the pulmonary valve replacement had a significantly lower risk of in-hospital mortality risk compared to Aorta, Mitral, and tricuspid valves replacement. The risk of mortality was decreased in both continues and individual suture types, compared to suture-less type, the risk of mortality was decreased by 0.052 (95% CI: 0.015, 0. 16), and 0.082 (95% CI: 0.023, 0.28) for Continues and Individual suture types, respectively (Table 3).

Table 3. The predictors of in-hospital mortality: Results of final multiple Bayesian logistic regression model in multivariate statistical analysis.

variable	Posterior Odds ratio	SE	Adjusted P	95% CI for OR
Suture Type	-	-	-	-
Individual/suture less	0.082	0.640	0.001	[0.023, 0.289]
Continues/suture less	0.052	0.606	0.001	[0.015, 0.16]
CPB time	1.010	0.003	0.010	[1.003, 1.017]
BUN-end	1.102	0.016	0.001	[1.070, 1.138]

variable	Posterior Odds ratio	SE	Adjusted P	95% CI for OR
INR After day4	1.339	0.157	0.066	[0.987, 1.807]
PTT	1.025	0.007	0.005	[1.012, 1.039]
Hemoglobin-end	0.606	0.154	0.001	[0.437, 0.813]
Platelet Count-end	0.973	0.004	0.001	[0.965, 0.981]
Creatinine-end	0.618	0.196	0.015	[0.425, 0.903]
Na-end	1.173	0.035	0.001	[1.096, 1.262]
K-end	7.918	0.310	0.001	[4.438, 15.359]
Proteinuria	-	-	-	-
1+/0	0.964	0.648	0.738	[0.280, 3.437]
2+/0	9.928	0.614	0.003	[2.850, 33.109]
Replaced valve site	-	-	-	-
Aorta, Mitral or both/ pulmonary	14.204	1.316	0.079	[1.382, 210.84]
Tricuspid/ pulmonary	46.754	1.487	0.016	[3.323, 1088.657]

The abbreviations are similar to Table 2.

The results of ROC curve showed that the last BUN, Cr, Hb, potassium, Plt count, and CPB time had the highest predictive values (Opt cut-off: 27, AUC: 0.862, Se: 67.91%, Sp: 89.83%, P<0.001; Opt cut-off: 1.1, AUC: 0.841, Se: 74.14%, Sp: 81.81%, P<0.001; Opt cut-off: 8.8, AUC: 0.830, Se: 68.32%, Sp: 86.50%, P<0.001; Opt cut-off: 4.6, AUC: 0.840, Se: 71.74%, Sp: 89.78%, P<0.001; Opt cut-off: 140, AUC: 0.862, Se: 69.78%, Sp: 90.67%, P<0.001; Opt cut-off: 139, AUC: 0.703, Se: 55.28%, Sp: 75.67%, P<0.001, Respectively) to predict in-hospital mortality, respectively.

Prediction models

Feature importance for pre-operative, intra-operative, post-operative and all features datasets are presented in Figure 2, Panel A-D, respectively.

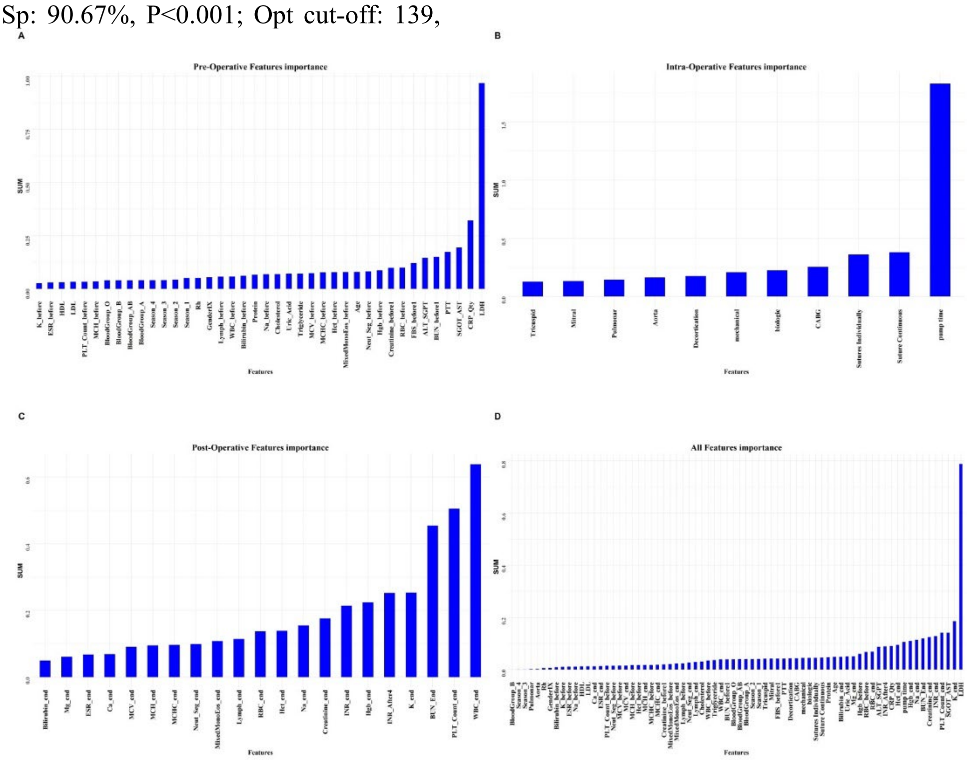

Figure 2. Feature importance for pre-operative, intra-operative, post-operative and all features datasets. Abbreviations are similar to Table 2 and Figure 1.

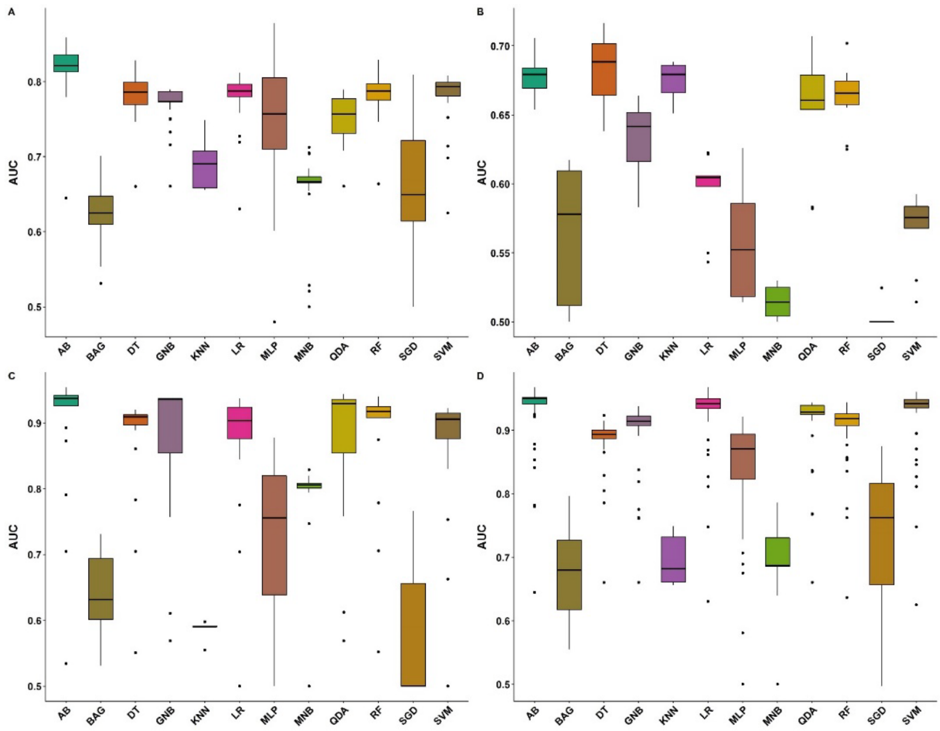

Figure 3. **AUC box plots of different machine learning classifiers among the different number of features in 4 sub-datasets [Pre-operative (Panel A), intra operative (Panel B), post-operative (Panel C), and all features (Panel D)]. Abbreviations are similar to Table 1.**

According to Table 4 and Figure 3, AB classifier had better performance than other classifier in all datasets (AUC: 0.82 ± 0.033, 0.68 ± 0.014, 0.90 ± 0.100, and 0.93 ± 0.049 for pre-operative, intra-operative, post-operative, and all features' datasets, respectively). On the other hand, BNB classifier had the lowest performance in almost all datasets (AUC: 0.500 ± 0.000, 0.52 ± 0.015, 0.50 ± 0.000, and 0.52 ± 0.020 for pre-operative, intra-operative, post-operative, and all features' datasets, respectively).

Table 4. **AUC of different machine learning classifiers among the different number of features in 4 sub-datasets**

Classifier	Pre-Operative	Intra-Operative	Post-Operative	All
Adaptive boosting (AB)	0.82 ± 0.033	0.68 ± 0.014	0.90 ± 0.100	0.93 ± 0.049
Bagging (BAG)	0.62 ± 0.034	0.56 ± 0.050	0.64 ± 0.055	0.67 ± 0.064
Bernoulli naïve bayes (BNB)	0.50 ± 0.000	0.52 ± 0.015	0.50 ± 0.000	0.52 ± 0.020
Decision trees (DT)	0.78 ± 0.029	0.68 ± 0.024	0.87 ± 0.090	0.89 ± 0.034
Gaussian naive bayes (GNB)	0.77 ± 0.023	0.63 ± 0.030	0.88 ± 0.110	0.90 ± 0.045
k-nearest neighbors (KNN)	0.69 ± 0.027	0.68 ± 0.014	0.59 ± 0.009	0.70 ± 0.038
Logistic regression (LR)	0.78 ± 0.031	0.60 ± 0.026	0.87 ± 0.100	0.93 ± 0.055
Multi-layer Perceptron (MLP)	0.75 ± 0.076	0.56 ± 0.041	0.71 ± 0.130	0.84 ± 0.077
Multinomial naïve Bayes (MNB)	0.66 ± 0.044	0.51 ± 0.012	0.79 ± 0.068	0.70 ± 0.039
Quadratic discriminant analysis (QDA)	0.75 ± 0.027	0.66 ± 0.041	0.88 ± 0.11	0.92 ± 0.048
Random forest (RF)	0.78 ± 0.028	0.66 ± 0.022	0.88 ± 0.095	0.91 ± 0.045
Stochastic gradient descent (SGD)	0.66 ± 0.084	0.5 ± 0.0074	0.56 ± 0.09	0.73 ± 0.097
Support vector machines (SVM)	0.78 ± 0.034	0.57 ± 0.024	0.86 ± 0.1	0.93 ± 0.056

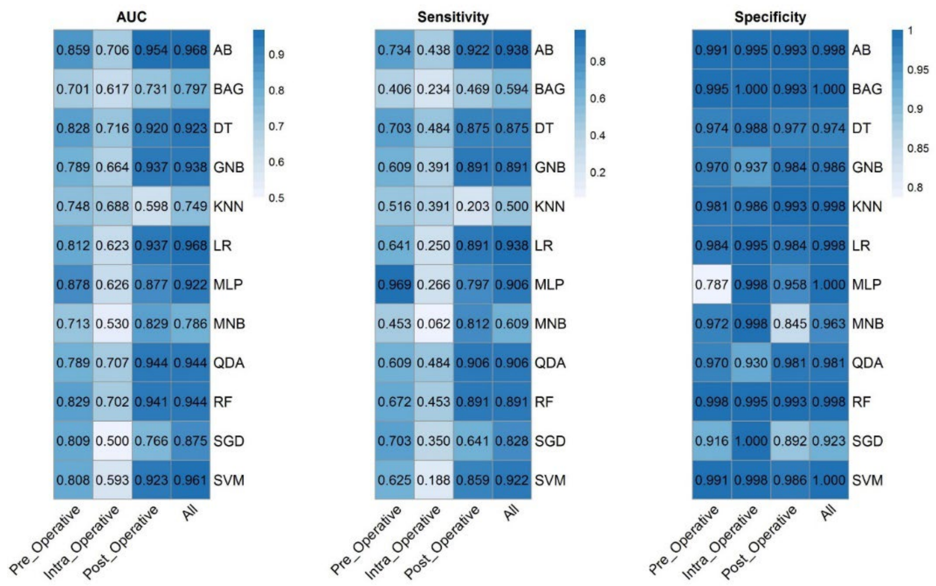

Figure 4. **Maximum AUC for the different classifiers in different numbers of features for 4 sub-datasets along with the sensitivity and specificity based one the maximum AUC. Abbreviations are similar to Table 1.**

Figure 4 indicates maximum AUC for different classifier in different number of features for 4 sub datasets. In pre-operative dataset, MLP, AB and DT with AUC: 0.88, 0.86 and 0.83 had highest AUC max, respectively. In intra-operative, DT, AB and QDA with AUC: 0.72, 0.71, 0.71 had highest AUC max, respectively. In post-operative, AB had AUC: 0.95 and followed by LR, RF, GNB and QDA had AUC: 0.94. The last dataset with all features, AB and LR had AUC: 0.97, followed by SVM with AUC: 0.96 and RF, QDA and GNB with AUC: 0.94.

In multivariate analysis performance of models was assessed by AUC, Sensitivity (SEN), Specificity (SPE), Positive predictive value (PPV), and Negative predictive value (NPV). Table 5 shows top models based on the above metric in pre-operative, intra-operative, post-operative, and all features' datasets. Each row indicates the top model based on that metric is bold.

Table 5. **Top prediction model of 4 datasets**

Dataset	NoF	Classifier	AUC	Sensitivity	Specificity	PPV	NPV
Pre-Operative	26	MLP	0.878	0.969	0.787	0.405	0.994
	26	MLP	0.878	0.969	0.787	0.405	0.994
	26	RF	0.780	0.563	0.998	0.973	0.938
	26	RF	0.780	0.563	0.998	0.973	0.938
	26	MLP	0.878	0.969	0.787	0.405	0.994
Intra-Operative	7	DT	0.716	0.484	0.948	0.585	0.925
	7	DT	0.716	0.484	0.948	0.585	0.925
	7	AB	0.663	0.328	0.998	0.955	0.908
	7	AB	0.663	0.328	0.998	0.955	0.908
	7	DT	0.716	0.484	0.948	0.585	0.925
Post-Operative	12	AB	0.954	0.922	0.986	0.908	0.988
	12	AB	0.954	0.922	0.986	0.908	0.988
	15	AB	0.936	0.875	0.998	0.982	0.982
	15	AB	0.936	0.875	0.998	0.982	0.982

Dataset	NoF	Classifier	AUC	Sensitivity	Specificity	PPV	NPV
	12	AB	0.954	0.922	0.986	0.908	0.988
	56	AB	0.968	0.938	0.998	0.984	0.991
	72	MLP	0.900	0.953	0.848	0.484	0.992
	56	AB	0.968	0.938	0.998	0.984	0.991
	56	AB	0.968	0.938	0.998	0.984	0.991
All	72	MLP	0.900	0.953	0.848	0.484	0.992

NOF: Number of features; PPV: Positive predictive value; NPV: Negative predictive value; MLP: Multi-layer Perceptron; RF: Random forest; DT: Decision tree; AB: Adaptive boost.

In pre-operative dataset MLP classifier with 26 features had the highest AUC (0.878) and the highest SEN (0.969) and NPV (0.994), but RF with 26 features had the highest SPE (0.998) and PPV (0.973). In intra-operative dataset, DT classifier with 7 features had the highest AUC (0.716) and the highest SEN (0.484) and NPV (0.925) but AB with 7 features had the highest SPE (0.998) and PPV (0.955). In post-operative dataset, AB classifier with 12 features had the highest AUC (0.954) and the highest SEN (0.922) and NPV (0.998) but AB with 15 features had the highest SPE (0.998) and PPV (0.982). In all features dataset, AB with 56 features had the highest AUC (0.968), SPE (0.998) and PPV (0.984) and MLP with 72 features had the highest SEN (0.953) and NPV (0.992) (Figure 5).

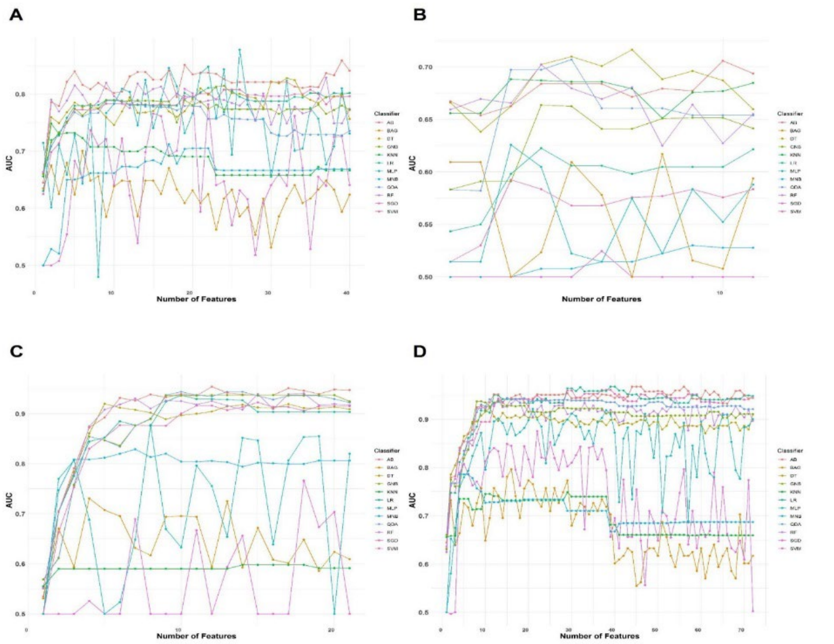

Figure 5. The impact of the number of features and performance of different model classifiers with AUC metric. Abbreviations are similar to Table 1.

Discussion

According to the results, the overall in-hospital mortality rate was 6.36%, which was consistent with the results reported by the Society of Thoracic Surgeons National Cardiac Surgery Database and National Inpatients Sample (NIS), which estimated the in-hospital mortality rate to be about 6.0% to 8.0% based on the target study population (12,13).

The prediction model results indicated that the AB classifier had the best performance in predicting in-hospital mortality after valve replacement surgery. AB is a well-known ensemble machine learning method for classification problems, which can effectively improve multiple weak predictive models by building an ensemble of models. AdaBoost focuses on the cases that were defectively predicted in the previous model, thereby ensuring that the more complex patterns are detected. In other words, AB yields its final output through the combination of the predictions of the individual models. AB offers the advantage of good generalization and can reduce both bias and variance (14). Generally, the AB model is reported to have an acceptable performance in the prediction of in-hospital mortality in patients admitted to ICU regardless of the cause of hospitalization **(15)**. However, the decision tree-based AB algorithms require a quality dataset and are sensitive to outlier and noisy data; thus, they have the potential to overfit the training set. AB had the worst performance in predicting survival after heart transplantation compared to RF and neural network ensembles **(16)**. Therefore, for complex and high dimension data, gradient boosting (GB) outperforms AB due to GB's system optimizations. Similarly, GB had the best performance (AUC: 0.78) for the prediction of acute kidney injury after cardiac surgery **(17)**. Furthermore, in a study by Kilic et al., GB was reported to be a promising model for prediction analytics in cardiac surgery mortality **(18)**. Hajianfar et al. also confirmed the previous results for the GB prediction model in methyl-guanine-DNA methyltransferase (MGMT) methylation status **(19)**. Compared to the other models, the BNB model had the worst performance overall in all four datasets. The obtained results may have been expected because the NB model unrealistically assumes that all of the features are independent and equally important **(20)**. Nilsson et al. used artificial neural network for the prediction of mortality after cardiac surgery, which exhibited an AUC of 0.81, while the MLP performance in our study had an AUC of 0.92 and was the better performance **(21)**. However, the study population in the investigation was comprised of all cardiac surgery procedures whereas our target population was limited to only those who underwent valvular replacement surgeries. In this study, preoperative LDH had the highest feature importance in the model design, which was consistent with the results of the study by Zhong et al. in which LDH and platelet had a higher rate of appearance among other predictors in patients undergoing open heart surgery **(22)**.

The performance of all models was relatively poor in the intra-operative dataset; however, machine learning in our study favored the inclusion of a wide range of variables that had acceptable performance in the pre- and post-operative datasets.

Limitations

This study had several limitations: 1) Only in-hospital mortality was considered and not the 30-day mortality, with all consequences correlated. The in-hospital mortality could be affected by a bad management of the discharge among other factors. 2) This was a single-center study, and the results may lack enough power to be generalized.

Conclusion

The results of machine learning algorithms in the prediction of in-hospital mortality were promising, and different algorithms performed better than the univariate results. Therefore, machine learning-based predictive models, such as AB, can be used to enhance significant data in prosthetic valve replacement surgery and to propose prediction and risk stratification models. AB is an advanced machine learning algorithm, which exhibited the highest discriminatory performance for the prediction of in-hospital mortality following valve replacement surgery.

Acknowledgments

We would like to express our thanks to Mr. Peyman Tabatabaei; Mrs. Mehrnoosh Ebrahimi; Mrs. Nasim Jafari for their fruitful consultations and cooperation through the study.

Conflict of Interest

None Declared.

Funding

None Declared.

References

1. Kim MS, Cho SJ, Park SJ, et al. Frequency and clinical associating factors of valvular heart disease in asymptomatic Korean adults. Sci Rep. 2019;9(1):1-8. [DOI:10.1038/s41598-019-53277-0] [PMID] [PMCID]
2. Thonghong T, De Backer O, Søndergaard L. Comprehensive update on the new indications for transcatheter aortic valve replacement in the latest 2017 European guidelines for the management of valvular heart disease. Open Heart. 2018;5(1): e000753. [DOI:10.1136/openhrt-2017-000753] [PMID] [PMCID]
3. Birkmeyer NJ, Marrin CA, Morton JR, et al. Decreasing mortality for aortic and mitral valve surgery in Northern New England. Ann Thorac Surg. 2000;70(2):432-7. [DOI:10.1016/S0003-4975(00)01456-9] [PMID]
4. Goldstone AB, Chiu P, Baiocchi M, et al. Mechanical or biologic prostheses for aortic-valve and mitral-valve replacement. N Eng J Med. 2017; 377(19):1847-57. [PMID] [PMCID] [DOI:10.1056/NEJMoa1613792]
5. Nowicki ER, Birkmeyer NJ, Weintraub RW, et al. Multivariable prediction of in-hospital mortality associated with aortic and mitral valve surgery in Northern New England. Ann Thorac Surg. 2004; 77(6):1966-77. [DOI:10.1016/j.athoracsur.2003.12.035] [PMID]
6. Hermiller JB, Yakubov SJ, Reardon MJ, et al. Predicting early and late mortality after transcatheter aortic valve replacement. J Am Coll

Cardiol. 2016;68(4):343-52. [DOI:10.1016/j.jacc.2016.04.057] [PMID]

7. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Health J. 2019;6(2):94-8. [DOI:10.7861/futurehosp.6-2-94] [PMID] [PMCID]
8. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart. 2018;104(14):1156-64. [DOI:10.1136/heartjnl-2017-311198] [PMID]
9. Arsang-Jang S, Kelishadi R, Motlagh ME, Heshmat R, Mansourian M. Temporal trend of non-invasive method capacity for early detection of metabolic syndrome in children and adolescents: A Bayesian multilevel analysis of pseudo-panel data. Ann Nutr Metab. 2019; 75(1):55-65. [DOI:10.1159/000500274] [PMID]
10. Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. J Stat Softw. 2017;80(1):1-28. [DOI:10.18637/jss.v080.i01]
11. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011; 12:2825-30.
12. Isaacs AJ, Shuhaiber J, Salemi A, Isom OW, Sedrakyan A. National trends in utilization and in-hospital outcomes of mechanical versus bioprosthetic aortic valve replacements. J Torac Cardiovasc Surg. 2015; 149(5):1262-9, e3. [DOI:10.1016/j.jtcvs.2015.01.052] [PMID]
13. Astor BC, Kaczmarek RG, Hefflin B, Daley WR. Mortality after aortic valve replacement: results from a nationally representative database. Ann Thorac Surg. 2000; 70 (6):1939-45. [DOI:10.1016/S0003-4975(00)01670-2] [PMID]
14. Schapire RE. Explaining adaboost. Empirical inference: Springer; 2013. 37-52. [DOI:10.1007/978-3-642-41136-6_5]
15. Stenwig E, Salvi G, Rossi PS, Skjærvold NK. Comparative analysis of explainable machine learning prediction models for hospital mortality. BMC Med Res Methodol. 2022; 22(1):1-14. [DOI:10.1186/s12874-022-01540-w] [PMID] [PMCID]
16. Ayers B, Sandholm T, Gosev I, Prasad S, Kilic A. Using machine learning to improve survival prediction after heart transplantation. J Card Surg. 2021; 36(11):4113-20. [DOI:10.1111/jocs.15917] [PMID]
17. Lee HC, Yoon HK, Nam K. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. J Clin Med. 2018; 7(10):322. [DOI:10.3390/jcm7100322] [PMID] [PMCID]
18. Kilic A, Goyal A, Miller JK, et al. Predictive utility of a machine learning algorithm in estimating mortality risk in cardiac surgery. Ann Thorac Surg. 2020; 109(6):1811-9. [DOI:10.1016/j.athoracsur.2019.09.049] [PMID]
19. Hajianfar G, Shiri I, Maleki H, et al. Noninvasive O6 methylguanine-DNA methyltransferase status prediction in glioblastoma multiforme cancer using magnetic resonance imaging radiomics features: univariate and multivariate radiogenomics analysis. World Neurosurg. 2019; 132: e140-e61. [DOI:10.1016/j.wneu.2019.08.232] [PMID]
20. Mansour NA, Saleh AI, Badawy M, Ali HA. Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy. J Ambient Intell Humaniz Comput. 2021:1-33. [DOI:10.1007/s12652-020-02883-2] [PMID] [PMCID]
21. Nilsson J, Ohlsson M, Thulin L, Höglund P, Nashef SA, Brandt J. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. J Thorac Cardiovasc Surg. 2006; 132(1):12-9. [DOI:10.1016/j.jtcvs.2005.12.055] [PMID]
22. Zhong Z, Yuan X, Liu S, Yang Y, Liu F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. Sci Rep. 2021; 11(1):1-10. [PMID] [PMCID] [DOI:10.1038/s41598-021-83020-7]

How to Cite This Article:

Shojaeifard M, Ahangar H, Gohari S, Oveisi M, Maleki M, Reshadmanesh T, et al. Assessment of Machine Learning Approaches to Predict in-Hospital Mortality in Patients Underwent Prosthetic Heart Valve Replacement Surgery. J Adv Med Biomed Res. 2023; 31(146): 210-20.