Machine learning vs. conventional methods for prediction of 30-day readmission following percutaneous mitral edge-to-edge repair

Samian Sulaiman[a,*], Akram Kawsara[a], Abdallah El Sabbagh[b], Abdulah Amer Mahayni[c], Rajiv Gulati[c], Charanjit S. Rihal[c], Mohamad Alkhouli[c]

[a]Division of Cardiology, West Virginia University, Morgantown, WV, United States of America

[b]Department of Cardiovascular Disease, Mayo Clinic, Jacksonville, FL, United States of America

[c]Department of Cardiovascular Disease, Mayo Clinic, Rochester, MN, United States of America

Abstract

Background: Identifying predictors of readmissions after mitral valve transcatheter edge-to-edge repair (MV-TEER) is essential for risk stratification and optimization of clinical outcomes.

Aims: We investigated the performance of machine learning [ML] algorithms vs. logistic regression in predicting readmissions after MV-TEER.

Methods: We utilized the National-Readmission-Database to identify patients who underwent MV-TEER between 2015 and 2018. The database was randomly split into training (70 %) and testing (30 %) sets. Lasso regression was used to remove non-informative variables and rank informative ones. The top 50 informative predictors were tested using 4 ML models: ML-logistic regression [LR], Naive Bayes [NB], random forest [RF], and artificial neural network [ANN]/For comparison, we used a traditional statistical method (principal component analysis logistic regression PCA-LR).

Results: A total of 9425 index hospitalizations for MV-TEER were included. Overall, the 30-day readmission rate was 14.6 %, and heart failure was the most common cause of readmission (32 %). The readmission cohort had a higher burden of comorbidities (median Elixhauser score 5 vs. 3) and frailty score (3.7 vs. 2.9), longer hospital stays (3 vs. 2 days), and higher rates of non-home

---

*Corresponding author at: Department of Cardiovascular Medicine, West Virginia University School of Medicine, 1 Medical Center Dr, Morgantown, WV 26506, United States of America. samian.sulaiman@hsc.wvu.edu (S. Sulaiman).

discharges (17.4 % vs. 8.5 %). The traditional PCA-LR model yielded a modest predictive value (area under the curve [AUC] 0.615 [0.587–0.644]). Two ML algorithms demonstrated superior performance than the traditional PCA-LR model; ML-LR (AUC 0.692 [0.667–0.717]), and NB (AUC 0.724 [0.700–0.748]). RF (AUC 0.62 [0.592–0.677]) and ANN (0.65 [0.623–0.677]) had modest performance.

Conclusion: Machine learning algorithms may provide a useful tool for predicting readmissions after MV-TEER using administrative databases.

Keywords

Transcatheter edge-to-edge repair; Machine learning; National Readmission Database

Introduction/Background

Mitral valve transcatheter edge-to-edge repair (MV-TEER) is an effective alternative to surgery in patients with severe symptomatic degenerative mitral regurgitation. MV-TEER is also recommended as a first-line intervention for patients with severe functional MR who remain symptomatic despite maximum guideline-directed medical therapy [1]. However, readmissions after MV-TEER remain common. In 2018, readmission rates after MV-TEER in the US were 14.7 %, 28.1 %, and 37.4 % at 30-, 90-, and 180- days [2]. Attempts have been made to identify predictors of readmissions after MV-TEER, however, those attempts were limited to small single-center studies [3,4] or low-performance models [5].

On the other hand, there is a rising interest in utilizing machine learning [ML] to predict adverse events after cardiac interventions using large databases [6–10]. ML algorithms have achieved remarkable success in cardiac imaging and electrocardiography, where they were able to detect or classify subtle myocardial, valvular, or rhythm disorders [6,11–13]. This success is partially attributed to the enormous amount of data embedded in the raw echo images or electrocardiograms. Whether ML can achieve similar success in analyzing electronic medical records (EMR) remains a subject of ongoing debate. Despite the enthusiasm about exploring the predictive value of ML algorithms in EMR, concerns remain on whether such databases contain adequate 'granular data' to improve the discrimination of adverse events after an acute illness or a cardiac intervention. Hence, to date, studies reporting on the value of ML methods in 'big data' analyses in cardiology are sparse.

In this study, we explore whether ML algorithms can improve the prediction of 30-day readmission after MV-TEER compared with logistic regression, the most used conventional model.

1. Methods

1.1. Data source

We queried the National Readmission Database (NRD.) from October 1st, 2015, through December 31st, 2018. The NRD, part of the Healthcare Cost and Utilization Project (HCUP), is a nationwide sample that represents ~60 % of all US hospitalizations across 28 states. The NRD provides demographics, inpatient diagnoses and procedures, total costs,

primary payers, length of stay, and hospital characteristics. Additionally, it contains a patient linkage number that identifies discharges belonging to the same individual within the same state [14]. The NRD has been used extensively to study trends and outcomes of MV-TEER [2,15–17]. Because NRD is publicly available and de-identified, this study was deemed exempt by the Institutional Review Board.

1.2. Study population & endpoint

We identified hospital stays for adult patients (18 years) who underwent MV-TEER using validated the International Classification of Diseases 10th Clinical Modification (ICD-10-CM) codes (Table S1). We excluded patients transferred to another hospital, those who died, and those discharged in December or who had missing information (Fig. 1). The *Primary Endpoint* of the study was 30-day all-cause readmission after MV-TEER.

1.3. Testing and training cohorts

We randomly split the data into sets of training (70 %, 6595 hospitalizations) and testing (30 %, 2830 hospitalizations). We did not access the testing set after this stage to ensure "never-seen-before" data for testing. Because the training set was slightly imbalanced (15 % readmitted), we used bootstrap sampling with replacement to balance the training dataset. After oversampling, the training set included 11,220 hospitalizations (50 % readmitted). Class imbalance can illusively increase the classifier's performance in the training set [18]. In the training stage, the ML software splits the training set into training and validation sets to fine-tune the model's parameters before testing on the testing set.

1.4. Predictor's selection process

Our approach was a combination of human and machine contributions, as previously described [8]. First, we did an extensive literature review to identify potential predictors of readmission after MV-TEER. A prior study showed that the Elixhauser score performed the best in predicting 1-year survival after MV-TEER [19]. Therefore, we used the HCUP Elixhauser comorbidity software to build all Elixhauser variables [20]. Furthermore, we reviewed three readmission assessment tools [LACE index [21], HOSPITAL score [22], and the 8Ps screening tool [23]] and included corresponding predictors available in the NRD data. In addition, since frailty predicts outcomes post-cardiac procedures [24], we added a Hospital Frailty Risk Score to the model [25].

Finally, we used the HCUP Clinical Classification Software to build additional diagnoses and procedures variables [26]. HCUP CCS aggregates the entire ICD 10 diagnosis and procedure codes into 283 diagnosis and 223 procedure categories that span all body systems. The final pool of 561 variables allowed the machine to pick out the codes closely related to 30-day readmission rather than depending on our human judgment. Table S2 shows the complete list of variables.

1.5. Lasso regression for feature ranking

Irrelevant variables/features result in model overfitting in the training stage, negatively affecting its performance in the testing stage. Therefore, we used the lasso regression method with 10-fold cross-validation in the training dataset to remove non-informative

variables. Out of 561 included variables, lasso regression ranked only 332 variables in descending order of importance. We selected the top 50 predictors for our ML models (Table S3) [26].

1.6. Machine learning algorithms

Next, we used 4 commonly utilized ML algorithms: ML logistic regression [ML-LR], Naïve Bayes [NB], random forest [RF], and artificial neural network [ANN]. ML-LR is a linear model whose output is an easy-to-interpret sum of variable values after applying coefficients to each variable. NB is a basic non-linear model that examines the probability of an outcome given the values of the variables (i.e., conditional probability). Rather than summing, NB multiplies probabilities. The independence of variables is a key assumption for NB and could be unrealistic occasionally [27,28]. RF is a modestly complex model that combines the output of multiple decision trees. ANN uses a network of interconnected units that links the variables to the outcome. Random forest and ANN appear as a "black box" because it is difficult to discern the exact calculations occurring in the model. We tested ML algorithm performance using the top 10/20/30/40/50 predictors ranked by lasso regression. We used Weka data mining software (version 3.8.5) for ML algorithms [29].

1.7. Classifier performance evaluation

We used the Receiver Operator Characteristic area under the curve (ROC AUC) as the main criteria, followed by precision-recall (PRC) AUC. We prioritized the ROC due to its wide use among researchers. While the ROC plots a tradeoff between sensitivity and (1-specificity), the PRC plots the precision and recall tradeoff. One major difference between ROC and PRC is that the PRC baseline moves with class distribution instead of the fixed baseline of 0.5 in ROC. Thus PRC might be a better evaluator for imbalanced data [30], such as ours, where the readmission rate is not close to 50 %. Additionally, we used other measures such as precision (i.e., positive predictive value), recall (i.e., sensitivity), and F-measure. The F score is the harmonic mean of precision and recall;

$$F = 2 \times \frac{precsion + recall}{precision + recall}.$$

1.8. Traditional statistical method

To compare ML performance with the traditional statistical methods, we performed a principal component analysis (PCA) followed by logistic regression. PCA decreases data dimensionality and is commonly used for outcome prediction [31]. It converted the 561 predictors to 166 components with an Eigenvalue > 1 (Fig. S1). To distinguish this from ML-LR, we used the term (PCA-LR) to describe traditional statistics LR. Next, we ran an LR analysis in the testing set, with the 30-day readmission variable as the outcome and the 166 components as predictors. Stata software was used for all traditional analyses [32].

Categorical variables were compared using the Chi-squared test and presented as percentages. Continuous variables with a normal distribution (e.g., age) were compared using a *t*-test and presented as means with standard deviation. Standardized differences in

means and proportions were calculated for continuous and categorical variables. We did not impute missing data because it represented <2 % of the entire data.

2. Results

2.1. Study cohort and baseline characteristics

A total of 10,659 index hospitalizations for MV-TEER (unweighted) were identified. After applying exclusion criteria, 9425 patients were included, of whome1,374 (14.6 %) were readmitted within 30 days (Fig. 1). Table 1 shows the differences in baseline characteristics between patients who were readmitted within 30 days and those who were not readmitted. In addition, baseline differences between the training and testing sets were small (standardized < 0.2), suggestive of appropriate randomization (Table 1). The 30-day readmission rate was 15 % in the training set and 13.7 % in the testing set. The most common causes of readmission were heart failure (32 %), cardiac dysrhythmia (6.5 %), and septicemia (5.8 %) (Table S4).

2.2. PCA-LR vs. ML classifiers' performance

In the testing set, the traditional PCA-LR model had a modest performance ROC AUC 0.615 (0.587–0.644) (Fig. 2). The predictive performance of ML algorithms varied between the models, with NB yielding the best predictive value (ROC AUC 0.724; PRC AUC 0.794) followed by ML-LR (ROC AUC 0.692; PRC AUC 0.76), ANN (ROC AUC 0.65; PRC AUC 0.747), and RF (ROC AUC 0.62; PRC AUC 0.732) (Table 2, Figs. 3, S2). The best performance of ML models was achieved utilizing 40 predictors in all but the RF model (Fig. 4, S Table S5).

3. Discussion

This study documents the utility of ML algorithms in building predictive models for readmissions after MV-TEER using administrative data. The predictive accuracy of ML algorithms in this study was superior to that of traditional risk prediction models, suggesting an additive value for ML in interventional cardiology practice. However, several key issues warrant further discussion.

The success of the ML algorithm depends on two key elements; the question being asked and the type of data being used to train the ML algorithms. Hernandez-Suarez et al. showed that an in-hospital mortality score developed using ML algorithms and an administrative database was superior to a commonly utilized TAVR risk score (TAVR in-hospital mortality score) developed using the TVT registry that contains extensive clinical data on all TAVRs performed in the US [8]. Khera et al. showed that XGBoost and meta-classifier ML models offered an improved resolution of mortality risk for high-risk individuals suffering an acute myocardial infarction [33]. Other investigators suggested the utility of ML in predicting the risk of post-TAVR bleeding or permanent pacemaker implantation using clinical databases [34,35].

To our knowledge, this is the first study that utilized a publicly available administrative database to compare the performance of ML with conventional statistical methods for the

prediction of readmission after MV-TEER. Our findings suggest that despite the limitations of administrative databases (e.g., the lack of echocardiographic and laboratory data), ML provided good predictive models for readmission (graphical abstract). As noted in the methodology section, the utilized ML algorithms ranged from simple (e.g., ML-LR) to more complex (e.g., ANN). Given the mathematical complexity of the ML models, it is difficult to determine why NB outperformed other ML models. However, we could speculate possible reasons. First, the underlying relationship between predictors and the outcome may be non-linear, and thus, an NB model (non-linear) explained it better than ML-LR (linear). Second, NB works well once the assumption of event independence is fulfilled. The Lasso regression has likely helped this assumption by removing correlated variables. Finally, RF and ANN are more complex models that have been shown to perform better with more extensive and complex data [36].

With traditional statistical methods, a large number of predictors of an event may hinder the accuracy and scalability of risk prediction models. Hence, the majority of risk scores include a limited number of predictors. With ML, although it is important to use certain techniques to rank the available variables according to their relevance to the studied event, typically, more variables are included in the models than with traditional analyses. Our findings suggest that the cutoff of 40 relevant variables was the most appropriate to achieve good discriminatory value in 3 out of the 4 ML models. Those variables included a mixture of baseline cardiovascular risk factors, non-cardiac comorbidities, procedure-related factors, and hospital characteristics. Although many of these variables are non-modifiable, identifying high-risk cohorts of patients for readmission might allow the consideration of additional measures that would mitigate their risk of readmission and potentially improve their clinical outcomes.

4. Limitations

Our study has several limitations. First, although the discriminatory power of the best ML model [NB] was good (AUC = 0.724), it may still be less than optimal for routine clinical application. However, it is still superior to other commonly utilized risk scores, such as the TVT/ ACC TAVR mortality score, which had a discrimination C statistic of 0.67 (95 % CI, 0.65–0.69) in the development group and 0.66 (95 % CI, 0.62–0.69) in the validation group [37]. Whether ML algorithms would perform even better when applied to the growing number of enhanced administrative datasets that contain laboratory and medication data remains to be seen. In addition, in our analysis, all 4 ML models had better PRC AUC (0.80 for NB) than ROC AUC. Some authors suggest that the PRC area might be more informative for binary classifier evaluation in imbalanced data. Hence, the predictive value of these models may even be better than what is suggested by the ROC AUC [30]. Second, the NRD is an administrative database hindered by the limitations of this type of database, such as coding errors and the lack of echocardiographic and laboratory data. Third, this study used one conventional method (logistic regression) because it is the most common. Therefore, our results should not be extended to other conventional models. However, other investigators' work supports our finding of low LR performance. For example, Ahuja et al. used LR to develop a predictive model of 30-day readmission after transcatheter MV repair; the c-statistic was 0.628, which is within our 95 % CI (0.587–0.644) [5]. Finally, although

MV-TEER is predominantly an inpatient procedure, same-day discharges have also been reported [38,39]. Therefore, some outpatient cases may not be captured in our study since the NRD includes only inpatient hospital stays.

On the other hand, because NRD is based on standard billing codes, any derived score can be implemented directly into hospital EMR across many health systems. The seamless integration into EMR is crucial because the risk calculators are suboptimally used by clinicians. Additionally, the risk variables are available before discharge, indicating that our risk model can identify high-risk patients before their discharge. Moving forward, the most crucial next step would be to evaluate targeted interventions such as follow-up calls, specifically targeting those identified as high-risk by our tool. Altogether, we believe that these data, acknowledging the limitation of the database, remain relevant to the practice. Furthermore, our findings may be viewed as hypothesis-generating to inspire further research to explore ML applications using widely available databases in cardiovascular medicine. Finally, future studies are needed to validate our models in specific patient populations.

5. Conclusions

ML algorithms may be helpful in forecasting readmission after MV-TEER using widely available administrative databases. Further studies are needed to explore the role of ML in the prediction of adverse events across a broad spectrum of cardiovascular interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations:

ANN	artificial neural network
LR	logistic regression
ML	machine learning
NB	Naïve Bayes
MV-TEER	mitral valve transcatheter edge-to-edge repair
AUC	area under the curve
PRC	precision-recall curve
RF	random forest
ROC	receiver operating characteristics
SD	standardized difference
D	day

| LOS | length of stay |

References

[1]. Writing Committee M, Otto CM, Nishimura RA, et al. 2020 ACC/AHA guideline for the management of patients with valvular heart disease: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 2021;77:450–500. [PubMed: 33342587]

[2]. Alkhouli M, Sulaiman S, Osman M, El Shaer A, Mayahni AA, Kawsara A. Trends in outcomes, cost, and readmissions of transcatheter edge to edge repair in the United States (2014–2018). Catheter Cardiovasc Interv 2021;99(3):949–55. [PubMed: 34520618]

[3]. Kessler M, Seeger J, Muche R, Wohrle J, Rottbauer W, Markovic S. Predictors of re-hospitalization after percutaneous edge-to-edge mitral valve repair by MitraClip implantation. Eur J Heart Fail 2019;21:182–92. [PubMed: 30178493]

[4]. Grasso C, Popolo Rubbio A, Capodanno D, et al. Incidence, timing, causes and predictors of early and late re-hospitalization in patients who underwent percutaneous mitral valve repair with the MitraClip system. Am J Cardiol 2018;121:1253–9. [PubMed: 29650238]

[6]. Ahuja KR, Nazir S, Ariss RW, et al. Derivation and validation of risk prediction model for 30-day readmissions following Transcatheter mitral valve repair. Curr Probl Cardiol 2021;48(3):101033. [PubMed: 34748783]

[7]. Lopez-Jimenez F, Attia Z, Arruda-Olson AM, et al. Artificial intelligence in cardiology: present and future. Mayo Clin Proc 2020;95:1015–39. [PubMed: 32370835]

[8]. Kagiyama N, Shrestha S, Farjo PD, Sengupta PP. Artificial intelligence: practical primer for clinical research in cardiovascular disease. J Am Heart Assoc 2019;8: e012788. [PubMed: 31450991]

[9]. Hernandez-Suarez DF, Kim Y, Villablanca P, et al. Machine learning prediction models for in-hospital mortality after transcatheter aortic valve replacement. JACC Cardiovasc Interv 2019;12:1328–38. [PubMed: 31320027]

[10]. Sardar P, Abbott JD, Kundu A, Aronow HD, Granada JF, Giri J. Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. JACC Cardiovasc Interv 2019;12:1293–303. [PubMed: 31320024]

[11]. AlHajji M, Alqahtani F, Alkhouli M. Contemporary trends in the utilization of administrative databases in cardiovascular research. Mayo Clin Proc 2019;94:1120–1. [PubMed: 31171122]

[12]. Kashou AH, Medina-Inojosa JR, Noseworthy PA, et al. Artificial intelligence-augmented electrocardiogram detection of left ventricular systolic dysfunction in the general population. Mayo Clin Proc 2021;96(10):2576–86. [PubMed: 34120755]

[13]. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet (Lond Engl) 2019; 394:861–7.

[14]. Casaclang-Verzosa G, Shrestha S, Khalil MJ, et al. Network tomography for understanding phenotypic presentations in aortic stenosis. J Am Coll Cardiol Img 2019; 12:236–48.

[15]. Healthcare Cost and Utilization Project (HCUP), editor. NRD overview Rockville, MD: Agency for Healthcare Research and Quality; April 2021 Retrieved 2 September 2021, from. https://www.hcup-us.ahrq.gov/nrdoverview.jsp.

[16]. Kawsara A, Alqahtani F, Rihal CS, Alkhouli M. Lack of association between the recommended annual volume thresholds for transcatheter mitral programs and safety outcomes of MitraClip implantation. JACC Cardiovasc Interv 2020;13:2822–4. [PubMed: 33303127]

[17]. Tripathi B, Sawant AC, Sharma P, et al. Short term outcomes after transcatheter mitral valve repair. Int J Cardiol 2021;327:163–9. [PubMed: 33278417]

[18]. Nazir S, Ahuja KR, Virk HUH, et al. Comparison of outcomes of transcatheter mitral valve repair (MitraClip) in patients <80 years versus >/=80 years. Am J Cardiol 2020;131:91–8. [PubMed: 32718547]

[19]. Chawla NV. Data mining for imbalanced datasets: an overview. In: Maimon O, Rokach L, editors. Data mining and knowledge discovery handbook Boston, MA: Springer US; 2010. p. 875–86.

[21]. Velu JF, Haas SD, Van Mourik MS, et al. Elixhauser comorbidity score is the best risk score in predicting survival after Mitraclip implantation. Struct Heart 2018;2:53–7.

[22]. Quality AfHRa. Tools archive for Elixhauser comorbidity software refined for ICD-10-CM Healthcare Cost and Utilization Project (HCUP); October 2020..

[23]. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. Can Med Assoc J 2010;182:551–7. [PubMed: 20194559]

[24]. Donzé JD, Williams MV, Robinson EJ, et al. International validity of the HOSPITAL score to predict 30-day potentially avoidable hospital readmissions. JAMA Intern Med 2016;176:496–502. [PubMed: 26954698]

[25]. The 8P Screening Tool, Accessed 8/21/2021.

[26]. Kim DH, Kim CA, Placide S, Lipsitz LA, Marcantonio ER. Preoperative frailty assessment and outcomes at 6 months or later in older adults undergoing cardiac surgical procedures: a systematic review. Ann Intern Med 2016;165:650–60. [PubMed: 27548070]

[27]. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. Lancet (Lond Engl) 2018;391:1775–82.

[28]. Healthcare Cost and Utilization Project (HCUP), editor. Tools archive for clinical classifications software refined Rockville, MD: Agency for Healthcare Research and Quality; March 2021www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccsr_archive.jsp.

[29]. Mueller J, Massaron L. Machine learning for dummies Hoboken, NJ: John Wiley & Sons, Inc.; 2016..

[30]. Witten IH, Frank E, Hall MA, Pal CJ. Data mining: practical machine learning tools and techniques 4th ed.. Cambridge, MA: Morgan Kaufmann; 2017..

[31]. Frank Eibe, Hall Mark A., and Witten Ian H. (2016). The WEKA workbench. Online appendix for "data mining: practical machine learning tools and techniques", Morgan Kaufmann, Fourth edition, 2016.

[32]. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015; 10:e0118432. [PubMed: 25738806]

[33]. Padró T, Cubedo J, Camino S, et al. Detrimental effect of hypercholesterolemia on high-density lipoprotein particle remodeling in pigs. J Am Coll Cardiol 2017;70: 165–78. [PubMed: 28683964]

[34]. StataCorp.. Stata statistical software: release 16. College Station, TX: StataCorp LLC; 2019..

[35]. Khera R, Haimovich J, Hurley NC, et al. Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiol 2021;6:633–41. [PubMed: 33688915]

[36]. Navarese EP, Zhang Z, Kubica J, et al. Development and validation of a practical model to identify patients at risk of bleeding after TAVR. JACC Cardiovasc Interv 2021;14:1196–206. [PubMed: 34112454]

[37]. Tsushima T, Al-Kindi S, Nadeem F, et al. Machine learning algorithms for prediction of permanent pacemaker implantation after transcatheter aortic valve replacement. Circ Arrhythm Electrophysiol 2021;14:e008941. [PubMed: 33685208]

[38]. Comparative Study on Classic Machine learning Algorithms|by Danny Varghese| Towards Data Science Retrieved 31 August 2021, from. https://towardsdatascience.com/comparative-study-on-classic-machine-learning-algorithms-24f9ff6ab222; 2021.

[39]. Edwards FH, Cohen DJ, O'Brien SM, et al. Development and validation of a risk prediction model for in-hospital mortality after transcatheter aortic valve replacement. JAMA Cardiol 2016;1:46–52. [PubMed: 27437653]

[40]. Chowdhury M, Buttar R, Rai D, et al. Same-day discharge after transcatheter mitral valve repair using MitraClip in a tertiary community hospital: a case series. Eur Heart J Case Rep 2021;5:ytab397. [PubMed: 34693199]

[41]. Marmagkiolis K, Kilic ID, Ates I, Kose G, Iliescu C, Cilingiroglu M. Feasibility of same-day discharge approach after transcatheter mitral valve repair procedures. J Invasive Cardiol 2021;33:E123–6. [PubMed: 33443488]
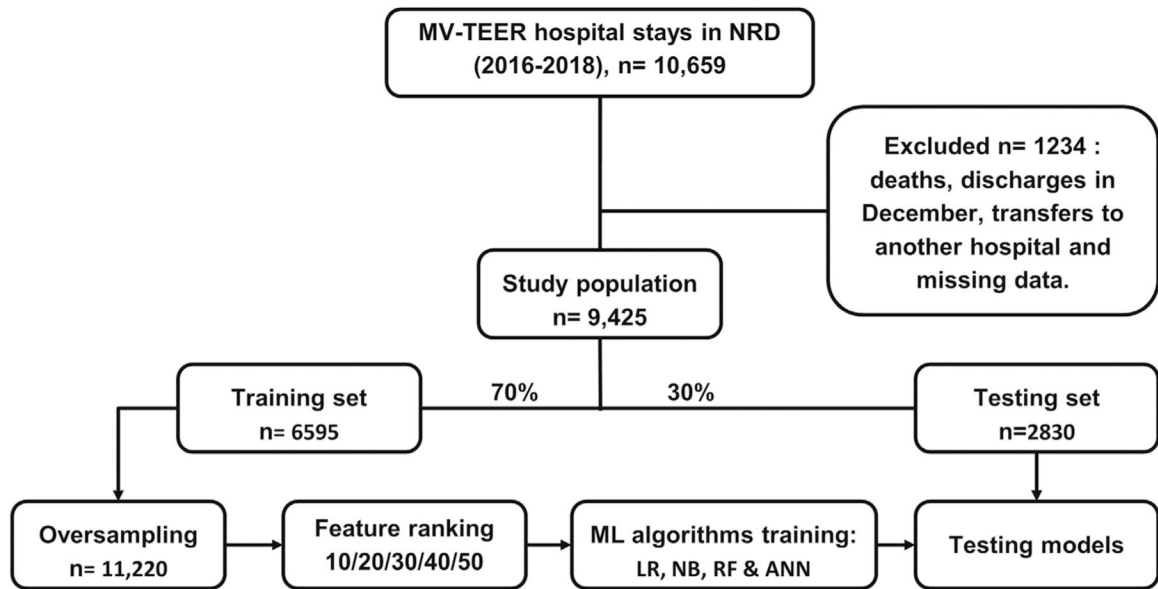
Fig. 1.
Study flow chart.

MV-TEER; mitral valve transcatheter edge-to-edge repair, ANN; Artificial neural network,

LR; Logistic regression, ML; Machine learning, NB; Naïve Bayes.
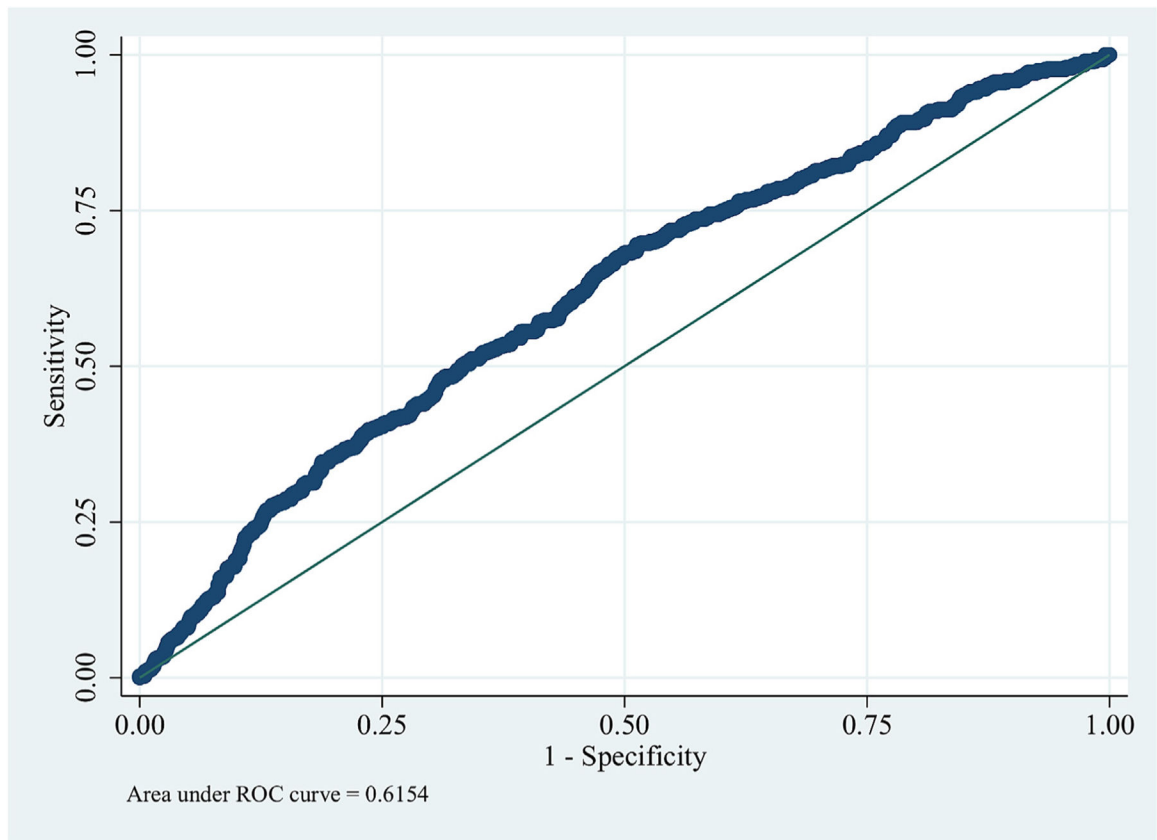
Fig. 2.
Prediction of readmission using traditional statistics (logistic regression).

ROC; receiver operator curve.
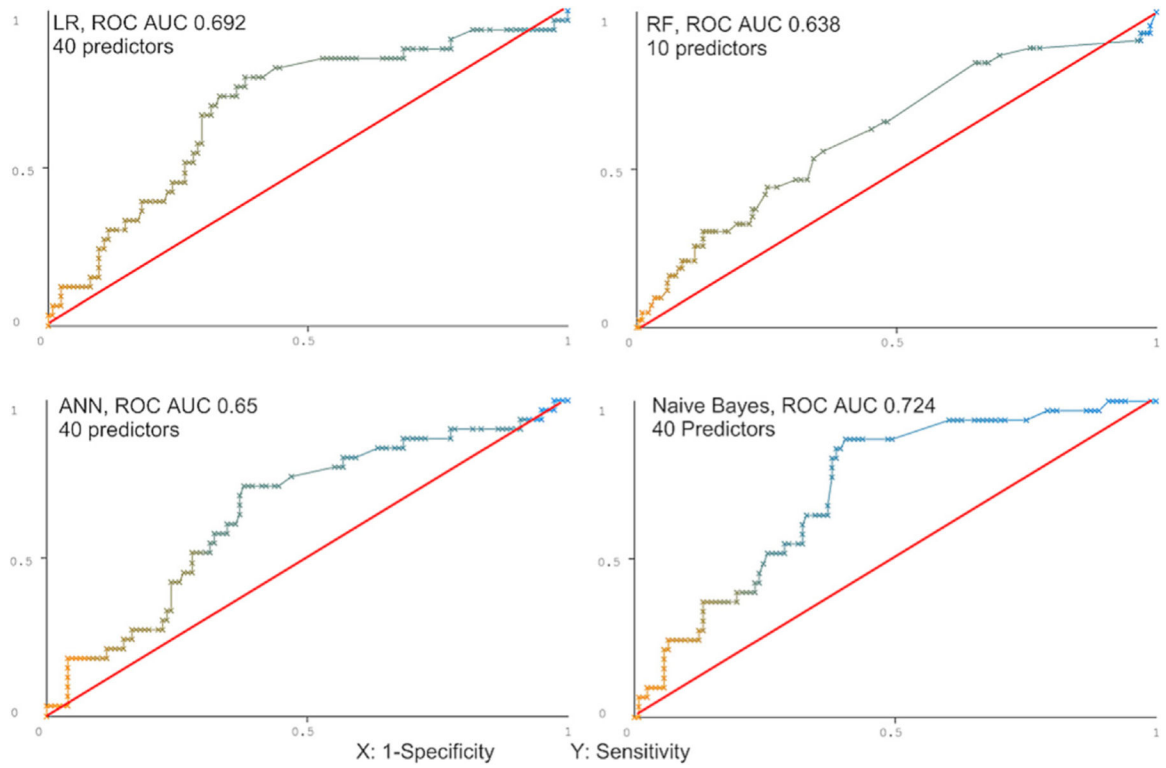
Fig. 3.

Prediction of readmission using machine learning algorithms.

ROC; receiver operator curve, ANN; Artificial neural network, LR; Logistic regression, ML; Machine learning, NB; Naïve Bayes.
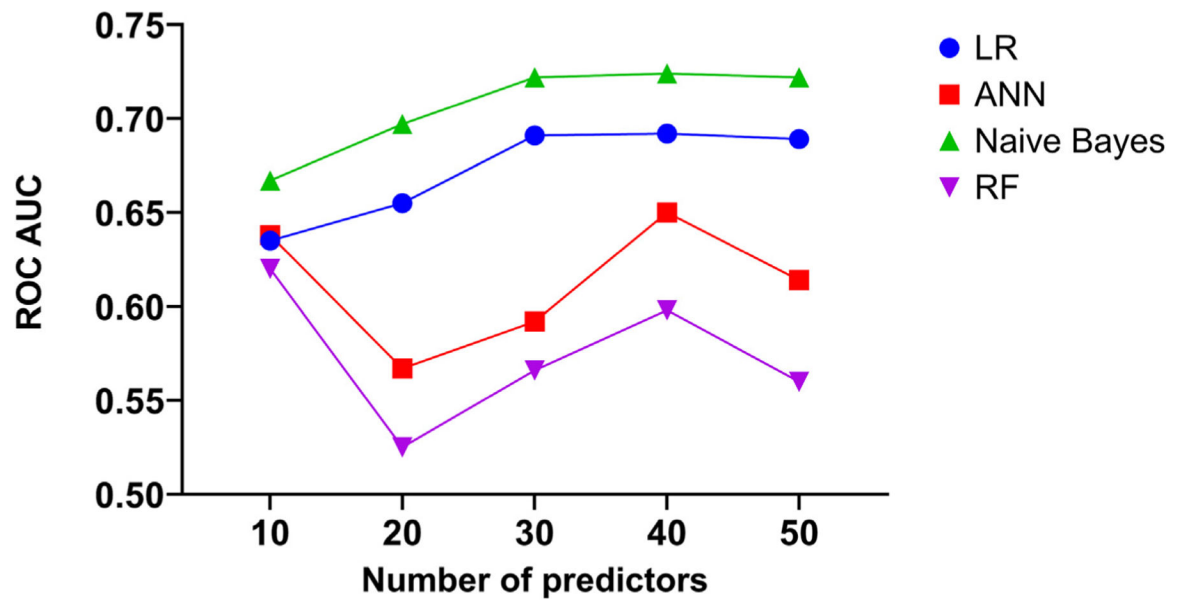
Fig. 4.

Performance of machine learning algorithms stratified by the number of variables included in the models.

ROC; receiver operator curve, ANN; Artificial neural network, LR; Logistic regression, ML; Machine learning, NB; Naïve Bayes.

Table 1

Comparison of baseline characteristics of the training and testing cohorts and per readmission status.

Baseline characteristics	Testing cohort (n = 2830)	Training cohort (n = 6595)	SD	No re-admit (n = 8051)	Re-admit (n = 1374)	SD
Demographics & hospital characteristics						
Age (mean ±SD), years	78.4 (10.0)	78.2 (10.1)	−0.02	78.3 (10.0)	78.5 (10.2)	0.02
Female sex	47.5 %	45.5 %	−0.04	45.3 %	50.7 %	0.11
Lowest quartile household income	19.8 %	20.5 %	0.02	20.2 %	21.0 %	0.02
Medicare insurance	90.4 %	90.1 %	−0.01	90.0 %	91.4 %	0.05
MV-TEER volume (>40 per year)	34.8 %	34.8 %	0.00	35.0 %	33.4 %	−0.03
Large bed size	74.8 %	76.6 %	0.04	76.3 %	74.3 %	−0.05
Teaching status	90.4 %	90.4 %	0.00	90.3 %	90.6 %	0.01
Urban location	94.8 %	94.1 %	−0.03	94.2 %	95.1 %	0.04
Clinical risk factors						
Cardiovascular co-morbidities						
Arrhythmia	70.2 %	69.3 %	−0.02	68.5 %	76.1 %	0.17
History of stroke	0.8 %	0.7 %	−0.01	0.7 %	1.2 %	0.05
Diabetes without complications	11.2 %	11.3 %	0.00	11.6 %	9.6 %	−0.06
Diabetes with complications	16.3 %	15.2 %	−0.03	14.5 %	21.3 %	0.19
Hypertension	30.7 %	31.4 %	0.01	31.4 %	29.8 %	−0.03
Myocardial infarction	2.0 %	2.3 %	0.02	2.1 %	3.0 %	0.06
Obesity	10.0 %	10.0 %	0.00	9.9 %	10.8 %	0.03
Vascular disease	14.0 %	13.6 %	−0.01	13.7 %	13.4 %	−0.01
Pulmonary circulation disorders	0.1 %	0.1 %	0.00	0.0 %	0.1 %	0.04
Valvular disease	4.5 %	4.5 %	0.00	4.3 %	5.3 %	0.05
Non-cardiovascular comorbidities						
Alcohol abuse	1.0 %	1.0 %	0.01	1.0 %	1.1 %	0.01
Blood loss anemia	0.8 %	0.8 %	0.00	0.7 %	1.2 %	0.05
Peptic ulcer disease	0.6 %	0.6 %	0.01	0.6 %	0.8 %	0.03
COPD	21.6 %	21.3 %	−0.01	20.4 %	27.3 %	0.17
Chronic pulmonary disease	27.3 %	26.7 %	−0.01	25.8 %	33.6 %	0.18
Coagulation disorder	9.4 %	9.4 %	0.00	9.1 %	11.4 %	0.08

Baseline characteristics	Testing cohort (n = 2830)	Training cohort (n = 6595)	SD	No re-admit (n = 8051)	Re-admit (n = 1374)	SD
Depression	6.5 %	7.5 %	0.04	7.0 %	8.3 %	0.05
Deficiency anemias	21.0 %	21.3 %	0.01	19.7 %	30.1 %	0.26
Elixhauser score	3.0 (0.0–8.0)	3.0 (0.0–7.0)	−0.02	3.0 (0.0–7.0)	5.0 (0.0–9.0)	−0.26
Median (IQR)						
GI bleeding	1.4 %	1.5 %	0.00	1.3 %	2.8 %	0.13
Hyponatremia	4.7 %	5.3 %	0.03	4.4 %	9.2 %	0.21
Hypothyroidism	17.9 %	18.6 %	0.02	18.0 %	20.4 %	0.06
Acquired immune deficiency	0.0 %	0.1 %	0.03	0.1 %	0.1 %	0.01
Lymphoma	0.7 %	1.0 %	0.03	0.9 %	0.9 %	0.00
Metastatic cancer	0.4 %	0.7 %	0.04	0.5 %	0.8 %	0.04
Liver disease	3.1 %	2.5 %	−0.03	2.6 %	3.2 %	0.04
Other neurological disorders	6.0 %	5.5 %	−0.02	5.6 %	6.3 %	0.03
Paralysis	2.1 %	1.7 %	−0.03	1.8 %	2.0 %	0.01
Psychoses	0.7 %	0.8 %	0.02	0.8 %	0.7 %	0.00
Pneumonia	3.2 %	3.1 %	−0.01	2.8 %	4.9 %	0.12
Collagen vascular diseases	3.4 %	3.9 %	0.03	3.7 %	4.0 %	0.01
Renal failure	38.2 %	37.7 %	−0.01	36.3 %	47.2 %	0.23
Solid tumor without metastasis	1.2 %	1.6 %	0.03	1.4 %	1.7 %	0.03
Septicemia	2.1 %	2.1 %	0.00	2.0 %	2.8 %	0.06
Weight loss	5.1 %	4.7 %	−0.02	4.5 %	6.9 %	0.11
Others						
Adverse drug events	3.1 %	2.7 %	−0.02	2.6 %	4.3 %	0.10
Discharge to skilled nursing facility	10.1 %	9.7 %	−0.02	8.5 %	17.4 %	0.30
Frailty score; Median (IQR)	2.9 (1.4–5.6)	2.9 (1.4–5.5)	0.01	2.9 (1.4–5.2)	3.7 (1.5–7.4)	0.31
Leaving against medical advice	0.0 %	0.1 %	0.02	0.0 %	0.2 %	0.07
Length of stay; Median (IQR)	2.0 (1.0–4.0)	2.0 (1.0–5.0)	−0.01	2.0 (1.0–4.0)	3.0 (1.0–9.0)	−0.31
Pre-admission within 30-days	18.1 %	17.3 %	−0.02	16.0 %	26.6 %	0.28
Therapeutic errors	0.2 %	0.1 %	−0.03	0.1 %	0.2 %	0.03

IQR; interquartile range, N; number, SD; standard deviation.

Table 2

Classifiers' predictive performance in the testing set.

Classifier	Precision	Recall	F-Measure	ROC area	PRC area
LR (40 predictors)	0.775	0.688	0.714	0.692 (0.667–0.717)	0.76 (0.738–0.782)
Naïve Bayes (40 predictors)	0.72	0.701	0.71	0.724 (0.700–0.748)	0.794 (0.774–0.814)
ANN (40 predictors)	0.73	0.662	0.686	0.65 (0.623–0.677)	0.747 (0.725–0.769)
RF (10 predictors)	0.721	0.638	0.668	0.62 (0.592–0.677)	0.732 (0.709–0.769)

ANN; artificial neural network, LR; logistic regression, PRC; precision-recall, RF; random forest, ROC; receiver operator characteristics.