Machine Learning Prediction Models for In-Hospital Mortality After Transcatheter Aortic Valve Replacement

Dagmar F. Hernandez-Suarez, MD, MSc[1], **Yeunjung Kim, MD, MPH**[2], **Pedro Villablanca, MD, MSc**[3], **Tanush Gupta, MD**[4], **Jose Wiley, MD, MPH**[4], **Brenda G. Nieves-Rodriguez, BSc**[5], **Jovaniel Rodriguez-Maldonado, BSc**[5], **Roberto Feliu Maldonado, BSc**[5], **Istoni da Luz Sant'Ana, PhD**[5], **Cristina Sanina, MD**[4], **Pedro Cox-Alomar, MD, MPH**[6], **Harish Ramakrishna, MD**[7], **Angel Lopez-Candales, MD**[1], **William W. O'Neill, MD**[3], **Duane S. Pinto, MD, MPH**[8], **Azeem Latib, MD**[4], **Abiel Roche-Lima, PhD**[5]

[1]Division of Cardiovascular Medicine, Department of Medicine, University of Puerto Rico School of Medicine, San Juan, PR, USA

[2]Division of Cardiovascular Medicine, Department of Medicine, Yale University School of Medicine, New Haven, CT, USA

[3]Division of Cardiovascular Medicine, Department of Medicine, Henry Ford Hospital, Detroit, MI, USA

[4]Division of Cardiology, Department of Medicine, Montefiore Medical Center/Albert Einstein College of Medicine, New York, NY, USA

[5]Center for Collaborative Research in Health Disparities, University of Puerto Rico School of Medicine, San Juan, Puerto Rico

[6]Division of Cardiology, Department of Medicine, Louisiana State University, New Orleans, LA, USA

[7]Division of Cardiovascular and Thoracic Anesthesiology, Mayo Clinic, Phoenix, AZ, USA

[8]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

Abstract

Objectives: To develop and compare an array of Machine Learning (ML) methods to predict in-hospital mortality after TAVR in the United States (US).

Address for Correspondence: Dagmar F. Hernandez-Suarez, MD, Division of Cardiovascular Medicine., University of Puerto Rico School of Medicine., PO Box 365067. San Juan, Puerto Rico 00936-5067., Tel: (787)758-2525; Fax: (787)754-1739., dagmar.hernandez@upr.edu.

Background: Existing risk prediction tools for in-hospital complications in patients undergoing transcatheter aortic valve replacement (TAVR) have been designed using statistical modeling approaches and have certain limitations.

Methods: Patient data was obtained from the National Inpatient Sample (NIS) database from 2012–2015. The data was randomly divided into a development cohort (n=7,615) and a validation cohort (n=3,268). Logistic Regression (LR), Artificial Neural Network (ANN), Naïve Bayes (NB) and Random Forest (RF) ML algorithms were applied to obtain in-hospital mortality prediction models.

Results: A total of 10,883 TAVRs were analyzed in our study. The overall in-hospital mortality was 3.6%. Overall, prediction models' performance measured by Area Under the Curve (AUC) were good (> 0.80). The best model was obtained by LR, AUC=0.92 (95% CI, 0.89–0.95). Most obtained models plateaued after introducing 10 variables. Acute kidney injury was the main predictor of in-hospital mortality ranked with the highest mean importance in all the models. The NIS TAVR score showed the best discrimination among available TAVR prediction scores.

Conclusions: Machine learning methods can generate robust models to predict in-hospital mortality for TAVR. The NIS TAVR score should be considered for prognosis and shared decision making in TAVR patients.

CONDENSED ABSTRACT

Prediction tools for in-hospital mortality after transcatheter aortic valve replacement (TAVR) are limited. Machine learning (ML) application on TAVR-big data remains unknown. In this study we sought to develop and compare an array of ML methods to predict in-hospital mortality after TAVR in the United States. A total of 10,883 TAVRs were analyzed. Overall, prediction models' performance measured by Area Under the Curve (AUC) were good (> 0.80). The best model was obtained by Logistic Regression, AUC=0.92. Our study highlights the potential utility of artificial intelligence in generating robust models to predict adverse outcomes in TAVR patients.

Keywords

machine learning; mortality; transcatheter aortic valve replacement

INTRODUCTION

Transcatheter aortic valve replacement (TAVR) has recently emerged as the gold standard treatment for majority of the patients with severe symptomatic aortic stenosis (AS) [1]. Patients who are at high or prohibitive surgical risk have significant benefit over medical therapy and surgical aortic valve replacement (SAVR), respectively. Most recent evidence has expanded TAVR indication to include intermediate-risk patients [1,2], and the recent PARTNER 3 trial results suggest that TAVR will shortly become the preferred choice for the low-risk candidates [3]. Concerns related to periprocedural complications — particularly in-hospital mortality, are increasing as TAVR becomes widely used.

Prediction models for cardiovascular procedures are critical for patient selection, risk stratification, and to tailor therapy and determine prognosis. The American College of

Cardiology (ACC) and the Society of Thoracic Surgeons (STS) have recently developed a model intended to predict the risk of in-hospital death based on the STS/ACC Transcatheter Valve Therapy (TVT) Registry data [4,5]. Other TAVR-specific risk models have been also produced in recent years [6–12]. However, these models have only relied on conventional statistical methods which carry inherent limitations that might affect its application and performance in large data sets with multiple variables and samples [13].

Machine learning (ML) is a discipline of computer science that focuses on predicting outcomes of complex datasets using algorithms that iteratively learn from data [13,14]. Rather than considering fixed assumptions on data behavior and variable preselection, ML algorithms allows the data to create the model by detecting or learning underlying patterns [15]. This concept has been already introduced in cardiovascular studies to obtain predictor models for adverse cardiovascular events [15,16]. They are yet to be applied in data derived from TAVR patients. We therefore aimed to compare an array of ML algorithms to predict in-hospital mortality after TAVR in the United States (US).

METHODS

2.1. Data source

Data was obtained from the National Inpatient Sample (NIS) database files. The NIS is a publicly available de-identified database of hospital inpatient stays, sponsored by the Agency for Healthcare Research and Quality as part of the Healthcare Cost and Utilization Project [17]. Further description of the NIS database has been previously published [18,19].

The International Classification of Diseases, Ninth Edition, Clinical Modification (ICD-9-CM) codes were used to identify all patients aged >18 years undergoing TAVR (codes 35.05 and 35.06) between January 1, 2012 and September 30, 2015. This last cutoff date was chosen because the new implementation of the ICD-10-CM took efect in October 1, 2015. Records with missing data on age, gender, or in-hospital mortality were excluded from analysis (n=10). The final study population (n= 10,883) was allocated into two groups: 1) Patients that survived hospitalization (Alive, n= 10,493) and 2) patients who died (Deceased n= 390). A list of ICD-9-CM codes used for selection of the study population is provided in Table 1S in the Data Supplement. Baseline characteristics were obtained either using the Elixhauser's comorbidities [20] or the corresponding ICD-9-CM codes described in Table 1S of the Data Supplement. The main endpoint of our study was all-cause in-hospital mortality. Unfortunately, other VARC-2 definitions are not well established in the NIS database and therefore, in-hospital complications were merely obtained using the ICD-9-CM coding system. A similar approach has been previously implemented by other authors [19, 21].

2.2. Study Design

Four machine learning algorithms were applied to the TAVR cohort patients. Central Illustration shows the different methods for data extraction, preparation of training and test sets as well as development and testing of the ML algorithms. Initially, the data was extracted from the NIS database [17] and processed to obtain the patient cohort using a

tailored scripting developed in the programming language Python. The dataset was divided into training and testing sets for the development and validation of the ML algorithms, respectively. For this purpose, the study population was split into a 70% (7,342 alive and 273 deceased) of the negative and positive classes for the development cohort (training set) and 30% (3,151 alive and 117 deceased) of the negative and positive classes for the validation cohort (testing set). Since the training set was unbalanced for the negative and positive class values of the in-hospital mortality variable (i.e. 7,342 alive and 273 deceased, respectively) and the positive values were relatively rare, an unbalanced oversampling technique [22,23] was applied to obtain an equal representation for each value of the class. Consequently, the initial sample proportion of 96:4 (96% - 'alive' and 4% - 'deceased') was transformed to 50:50 (50% - 'alive' and 50% - 'deceased') to later train the models. The total sample obtained after oversampling was n=14,442.

A feature ranking method (see section 2.3) was used to determine variable or feature values intended to be more informative and non-redundant. Features were then considered in groups of 5/10/15/20/30/All (All = 43 variables, as represented in Table 1) organized by the ranks obtained after the feature ranking method. Machine learning algorithms integrated in WEKA (v. 3-7-12, New-Zealand) data mining software [24] were applied for each group of variables or features (i.e. 5/10/15/20/30/All). The resulting models were validated using the test set, and different precision metrics were computed, focused on the area under of the receiver operating curve (AUC) values.

2.3 Feature ranking

The process of extracting or ranking features is an excellent strategy to reduce dimensionality and the complexity of machine learning algorithms [25]. In this study, a regularized logistic regression with an L2 norm penalty as a feature scoring function was used to rank the variables in the cohort datasets [26]. The goal was to determine the influence that each variable had on the primary outcome. This method added the coefficients square magnitude as a penalty term to spread the range of possible values equally. Using the regularized logistic regression coefficients, the variables or features were ranked in descending order and subset by the highest 5/10/15/20/30 ranking variables values. All variables were also included. To perform this feature ranking process, Python 3.5 was used in conjunction with the scikit-learn v0.20.2 [27, 28].

2.4. Machine learning algorithms

Five top supervised-learning classification algorithms are regularly used as a starting point for binary classification problems in Machine Learning [29]. These algorithms are Logistic Regression (LR), Artificial Neural Networks (ANN), Naive Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM). All these algorithms were applied in this study; however, SVM was discarded because it failed to show a good fit for the dataset.

In all cases, the oversampled training sets for each set of variables or features (5/10/15/20/30/ALL) were used to develop the models. A separate test set (not included in the training of the models) was used to evaluate the models and compute the different accuracy metrics with a particular focus on the AUC values. For all cases we used the

implementation of the supervised-learning classification algorithms in WEKA software [24] for each group of variables or features. The parameters of these algorithms are illustrated in Table 2S.

Logistic Regression—Logistic regression is used to predict the probability of an event with a discrete dependent variable [30]. This model utilizes a sigmoid function to predict the logistic transformation of the probability for each class in the dependent variable. The logged odds classify the data points in a binary fashion. The lambda parameter used for the model was a ridge value of 1.0E-8 in addition to conjugate gradient descent. Conjugate gradient descent was applied to reduce the cost function in the model.

Artificial Neural Network—Inspired by biological networks, artificial neural networks are based on nodes and connections that loosely simulate the inner workings of a biological brain [31]. The model is learned based on multilayer perceptrons using a backpropagation algorithm [32], which is the standard algorithm for supervised-learning process. ANN models were obtained using the implementation of Weka software with the defaulted parameters. Parameters included the number of hidden layers as 'a', which means the total of variables plus the classes divided by 2, in our case (for ALL variables) it was (43+1)/2, for a total of 22 hidden layers.

Naïve Bayes—Based on Bayes theorem, naïve bayes is a probabilistic classifier with a strong assumption of independence among variables or features [33]. The results can be interpreted as a bayesian network, where nodes represent features that are connected to one another by directed arcs. The model outputs the probability that each feature belongs to a specific class independently and considers all of them in unison to determine the class. Similar to previous algorithms, Weka software was used to train NB algorithms with our training set. Default parameters that included the 'batchSize' for the preferred number of instances to process were used if the batch prediction was performed.

Random Forest—Random forest (RF) is a classification algorithm that applies an ensemble of decision trees and bootstrapping to sample training data and split branches in each tree [34]. The target in each split is to maximize the gained information from each random feature in each sample per tree. After evaluating the data points, the resulting class is the mode of the results of all trees. The algorithm was deployed in our dataset using the gini criterion for splitting with no pruning and an *int(log_2(#predictors) + 1)* number of features per split in each tree.

All these machine learning algorithms were compared using our dataset for each group of features (i.e. 5/10/15/20/30/all). Their performances were evaluated using the following metrics accuracy, precision, recall, f-score and area under of the receiver operating curve (AUC). We focused our decision for the best model on the AUC metric results. The obtained algorithm's parameters are illustrated in Table 2S.

2.5 Statistical Analyses

Continuous variables were compared using the 2-tailed Student's *t*-test while Chi-square or Fisher's exact tests were used for categorical data as appropriate. The performance of

machine learning's prediction algorithms was assessed by measuring the total AUC. Statistical analyses were performed with the use of R software, version 3.5.1 GUI 1.70 (R Core Team, 2018), and values of p <0.05 were considered statistically significant.

RESULTS

3.1 Baseline characteristics

Of the 10,883 adult patients who underwent TAVR between January 1, 2012 and September 30, 2015 in the NIS database, 390 (3.6%) died before being discharged. Demographic and hospital characteristics of the study population are depicted in Table 1. Patients who died were more likely to be older, female, with prior history of heart failure, chronic kidney disease (CKD), coagulopathy, other valvular diseases, pulmonary circulation disorders, fluid/electrolyte disorders, admitted to a hospital located in the Southern US, undergo TAVR via transapical approach as well as having more hospital complications such as stroke, myocardial infarction (MI), conversion to SAVR, cardiogenic shock, cardiac arrest, major bleeding, vascular complications, acute kidney injury (AKI) and sepsis. In addition, they were less likely to have history of smoking, prior coronary artery disease, prior myocardial infarction, carotid artery disease, diabetes mellitus, hypertension and to have undergone angiography, percutaneous coronary intervention and coronary artery bypass graft before the procedure. A total of 7,615 (70%) and 3,268 (30%) patients were randomly assigned to the development and validation cohorts respectively. Patient's characteristics were similar between the development and validation cohort with the exception of fluid and electrolytes disorders that were slightly increased in this latter.

3.2 Machine learning classifiers ' performance for in-hospital mortality.

From the four machine-learning algorithms applied to the NIS TAVR dataset patients, the model obtained by LR had the best discrimination (AUC=0.92; 95% CI, 0.89–0.95) (Figure 1). However, all models showed a good AUC for predicting in-hospital mortality [ANN (0.85; 95% CI, 0.82–0.88), NB (0.90; 95% CI, 0.88–0.92) and RF (0.90; 95% CI, 0.87–0.93). After including features iteratively, the performance in all ML algorithms increased up to the top 10 variables. Thereafter, models obtained by LR and NB had minimal performance's improvement while ANN and RF had a transient performance's decline (Table 3S) (Figure 2).

Each variable included in the study had varying importance over in-hospital mortality depending on the ML approach. Overall, AKI was the variable with greatest importance across all ML algorithms, followed by cardiogenic shock, fluid and electrolyte disorders, cardiac arrest, sepsis, dyslipidemia, hypertension, coagulopathy, current smoking, and vascular complications, respectively (Figure 3) (Online Table 4).

The prospective predictive performance of the studied models is illustrated in Table 2. Model's sensitivity, specificity, positive predictive value, negative predicted and f-score ranged from 84.8–96.3%, 16.7–83.9%, 95.0–96.5%, 13.7–83.8% and 0.90–0.96, respectively.

3.3 Characterization of available TAVR scores.

Established TAVR mortality prediction scores with AUC are represented in Table 3. Of note, the LR model obtained in our study (NIS TAVR) showed the best discrimination among all available in-hospital mortality (STS/ACC TVT [5]), 30-day mortality (FRANCE 2 [6], TARIS [7], OBSERVANT [8], CoreValve US [10] and UK TAVI [11]) and 1-year mortality (TAVI$_2$ [9], CoreValve US [10] and CAPRI [12]) TAVR prediction scores.

DISCUSSION

This study represents the first report using artificial intelligence (AI) in the development and validation of a risk prediction model for TAVR-related outcomes. AI has significant advantages for predicting events based on machine learning algorithms, and the statistical improvement with its application is not unique to this work [14–16]. Our results are compelling since the NIS TAVR score is excellent at predicting in-hospital mortality (AUC 0.92) with superior discrimination than other risk prediction models. With expanding indications for TAVR, a better predictor of periprocedural mortality is paramount for patient selection and prognosis. Our study is the second largest study to derive a risk prediction model in patients who underwent TAVR for management of aortic stenosis. The largest was the STS/ACC TVT score [5]. Both STS/ACC TVT and NIS TAVR scores look at in-hospital mortality as the primary outcome, and the covariates in each model include some aspects of patient comorbidity and the acuity of the hospitalization. STS/ACC TVT score includes a multivariable parameter (acuity status) to describe urgency of the procedure, the clinical status pre-procedure and cardiac arrest within 24 hours of the operation. Similarly, NIS TAVR score incorporates in-hospital cardiac arrest, shock, acute kidney injury, and sepsis into the score. While there are similarities, the main utility of the NIS TAVR score is to determine prognosis, due to the inclusion of post-procedural variables within the model. Given the differences in primary endpoints and that prior TAVR-related outcomes scores have been obtained using different populations, no statistical comparison can be performed among all the scores represented in Table 3. However, a first glance is enough to identify the outperformance of the NIS TAVR score, perhaps because of the inherent benefits of AI. Consequently, because the STS/ACC TVT registry has been designed with the VARC-2 definitions and has a more structured data, there may be an opportunity to improve this model using AI.

Although STS/ACC TVT is the largest study to date, wide adoption has not occurred in clinical practice since its publication in 2016 [5]. There are several reasons for its slow adoption. First, TAVR operators were largely committed to the procedure once the decision was made to offer TAVR during the multidisciplinary heart team. In general, the discussion focused around technical feasibility once a decision not undergo SAVR was made based on the STS risk score, frailty, and comorbidities. As a procedure of last resort, TAVR would be offered unless it was deemed prohibitive. Second, guidelines have continued to expand with more compelling data showing favorable outcomes for TAVR in different populations. In addition to the 2017 American Heart Association (AHA)/ACC focused update which added a class IIa recommendation for intermediate surgical risk patients, PARTNER 3 trial established the safety and effectiveness of TAVR in lower surgical risk patients as well [1,3].

These findings have led to the discussion regarding the appropriateness of SAVR as the first line therapy. Guidelines are likely to embrace broader recommendations supporting the superiority and durability of TAVR over SAVR [1,3]. Lastly, wide adoption may not have occurred because of the relative infancy of TAVR technology and ongoing exploration into long term durability and unknown predictors of clinical outcomes. Recent follow up studies show structural stability of TAVR as well as satisfactory clinical outcomes up to 10 years post-implantation; however, long-term clinical data remains insufficient at this time [35].

With the recent expansion to the lower surgical risk population, a TAVR risk predictor score has become critical for providers and patients for decision making. Our study highlights the potential of artificial intelligence to generate robust TAVR-related outcomes prediction tools. Machine learning for big data analysis has revolutionized the traditional way of conducting cardiovascular research. As a subset of AI, machine learning provides an innovative approach to data analysis and imaging interpretation beyond what is provided by conventional statistics [36]. The ability to automatically handle large multi-dimensional and multi-variety data could ultimately expose novel associations between specific features and important cardiovascular outcomes as well as identify trends and patterns that would not be apparent to investigators otherwise. In real life, this is akin to a taxi driver eagerly waiting at a busy street on Friday night which historically has been a safe spot for picking up passengers versus a driver receiving real time requests and the next client waiting at the end of the first trip. This ability to incorporate novel predictors from big data highlights the important applicability of AI in cardiovascular medicine. Rather than retrospectively developing models for risk prediction scores, real-time addition of newly discovered variable(s) can be feasible. As we continue to gather significant amount of patient data through the electronic medical record and increase the use of imaging in cardiology, AI is likely to become an essential tool for clinicians. Incorporation of AI into the expanding list of structural interventions will be a gift that keeps on giving as we are better able to predict the cutting edge of new innovations and interventions.

Limitations

NIS TAVR score is a risk prediction model that incorporates aspects of patient characteristics, hospital course, and the TAVR procedure. The main limitation of NIS dataset, by design, is the inability to restrict variables pre versus post procedure, which provides a dynamic nature to the NIS TAVR score. For example, using a variable such as cardiogenic shock may improve the prediction model when a cardiogenic shock occurs during the TAVR admission period. Yet, the use of covariates such as shock or arrest, are not exclusive to the NIS TAVR score and have been incorporated into different mortality risk scores for TAVR (Table 3) and other clinical scenarios [37]. Second, the NIS is an observational cohort using ICD-9 CM and procedure codes which relies on fidelity of hospital billing data and deriving clinical information from ICD-9 CM codes are an inherent limitation. Although procedures may be highly accurate for billing purposes, some non-related clinical diagnoses may be omitted and may not represent the true prevalence of risk factors. Third, since only patients undergoing TAVR between 2012–2015 were included in our study, our results may better represent the extreme and high surgical risk patient population. Thus, the NIS TAVR score application for the intermediate and low-risk patients

remains unknown. Nevertheless, since our score includes in-hospital complications, patient's surgical risk might not significantly affect its performance. Also, this score was not externally validated and therefore; future studies are necessary to better address the viability of this model. Finally, ML model complexity and difficult interpretation may hinder its reproducibility and application into the standard practice. However, given the growing amount of patient data - brought with the implementation of the Electronic Medical Record- as well as the rapid implementation of automated algorithms in other medical fields [38], we envision that the adoption of AI will shortly become inherent in clinical medicine.

CONCLUSIONS

Using a data meaning approach, we developed the NIS TAVR score to predict in-hospital mortality in TAVR patients. The good discrimination of this model reveals the potential of AI in the patient risk stratification process not just for TAVR but for any novel structural intervention. Further validation and application of ML into the day-to-day clinical practice is still warranted to better understand its true value in patients with severe aortic stenosis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

ABBREVIATIONS AND ACRONYMS

TAVR	transcatheter aortic valve replacement
AS	aortic stenosis
SAVR	surgical aortic valve replacement
ACC	American College of Cardiology
STS	Society of Thoracic Surgeons
TVT	transcatheter Valve Therapy
ML	machine learning
US	United States
NIS	National Inpatient Sample
ICD-9-CM	international classification of diseases, ninth edition, clinical modification
LR	logistic regression

ANN	artificial neural networks
NB	nai've bayes
RF	random forest
AUC	area under the receiver operating curve
CKD	chronic kidney disease
MI	myocardial infarction
AKI	acute kidney injury
AI	artificial intelligence
AHA	American Heart Association
PROM	predicted risk of mortality

REFERENCES

1. Nishimura RA, Otto CM, Bonow RO, et al. 2017 AHA/ACC Focused Update of the 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol. 2017;70(2):252–289. [PubMed: 28315732]

2. Chamandi C, Puri R, Rodriguez-Gabella T, Rodés-Cabau J. Latest generation transcatheter aortic valve replacement devices and procedures. Can J Cardiol 2017;33:1082–90. [PubMed: 28843322]

3. Mack MJ, Leon MB, Thourani VH, et al. Transcatheter Aortic-Valve Replacement with a Balloon-Expandable Valve in Low-Risk Patients. N Engl J Med 2019;380:1695–1705. [PubMed: 30883058]

4. TAVR In-Hospital Mortality Risk Calculator. https://tools.acc.org/tavrrisk/#!/content/evaluate/. Accessed February 28, 2018.

5. 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(1): 46–52. [PubMed: 27437653]

6. Iung B, Laouénan C, Himbert D, et al. Predictive factors of early mortality after transcatheter aortic valve implantation: individual risk assessment using a simple score. Heart 2014;100(13):1016–23. [PubMed: 24740804]

7. Seiffert M, Sinning JM, Meyer A, et al. Development of a risk score for outcome after transcatheter aortic valve implantation. Clin Res Cardiol 2014;103(8):631–40. [PubMed: 24643728]

8. Capodanno D, Barbanti M, Tamburino C, et al. A simple risk tool (the OBSERVANT score) for prediction of 30-day mortality after transcatheter aortic valve replacement. Am J Cardiol 2014;113(11):1851–8. [PubMed: 24837264]

9. Debonnaire P, Fusini L, Wolterbeek R, et al. Value of the "TAVI2-SCORe" Versus Surgical Risk Scores for Prediction of One Year Mortality in 511 Patients Who Underwent Transcatheter Aortic Valve Implantation. Am J Cardiol 2015;115(2):234–42. [PubMed: 25432413]

10. Hermiller JB Jr, Yakubov SJ, Reardon MJ, et al. Predicting early and late mortality after transcatheter aortic valve replacement. J Am Coll Cardiol 2016;68:343–352. [PubMed: 27443429]

11. Martin GP, Sperrin M, Ludman PF, et al. Novel United Kingdom prognostic model for 30-day mortality following transcatheter aortic valve implantation. Heart 2018;104(13): 1109–1116. [PubMed: 29217636]

12. Lantelme P, Eltchaninoff H, Rabilloud M, et al. Development of a Risk Score Based on Aortic Calcification to Predict 1-Year Mortality After Transcatheter Aortic Valve Replacement. JACC Cardiovasc Imaging 2019;12(1):123–132. [PubMed: 29778857]

13. Shouval R, Bondi O, Mishan H, Shimoni A, Unger R, Nagler A. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. Bone Marrow Transplant 2014;49(3):332–7. [PubMed: 24096823]

14. Deo RC. Machine Learning in Medicine. Circulation 2015;132(20):1920–30. [PubMed: 26572668]

15. Shouval R, Hadanny A, Shlomo N, et al. Machine learning for prediction of 30-day mortality after ST elevation myocardial infraction: An Acute Coronary Syndrome Israeli Survey data mining study. Int J Cardiol 2017;246:7–13. [PubMed: 28867023]

16. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. Circ Res 2017;121(9):1092–1101. [PubMed: 28794054]

17. Overview of the National (Nationwide) Inpatient Sample (NIS). https://www.hcup-us.ahrq.gov/nisoverview.isp. Accessed February 28, 2018.

18. Velagapudi P, Kolte D, Ather K, et al. Temporal Trends and Factors Associated With Prolonged Length of Stay in Patients With ST-Elevation Myocardial Infarction Undergoing Primary Percutaneous Coronary Intervention. Am J Cardiol 2018;122(2):185–191. [PubMed: 29751953]

19. Gupta T, Khera S, Kolte D, et al. Transcatheter Versus Surgical Aortic Valve Replacement in Patients With Prior Coronary Artery Bypass Grafting Trends in Utilization and Propensity-Matched Analysis of In-Hospital Outcomes. Circ Cardiovasc Interv 2018;11(4):e006179. [PubMed: 29643130]

20. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998;36(1):8–27. [PubMed: 9431328]

21. Gupta T, Goel K, Kolte D, et al. Association of Chronic Kidney Disease With In-Hospital Outcomes of Transcatheter Aortic Valve Replacement. JACC Cardiovasc Interv. 2017;10(20): 2050–2060.

22. Chawla NV. Data Mining for Imbalanced Datasets: An Overview In: Maimon O, Rokach L, editors. Data Mining and Knowledge Discovery Handbook. Boston, MA Springer, 2009: 875–886.

23. Ling C, Li C Data mining for direct marketing: Problems and solutions. KDD'98 Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining. 1998; 98: 73–79.

24. Witten IH, Frank E, Hall MA, Pal CJ. The WEKA Workbench. Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques. San Francisco, CA Morgan Kaufmann 2016.

25. Fakhraei Shobeir,a,b Hamid Soltanian-Zadeh,a,c and Farshad Fotouhid. Bias and Stability of Single Variable Classifiers for Feature Ranking and Selection. Expert Syst Appl. 2014; 14(15): 6945–6958. [PubMed: 25177107]

26. Feature selection, L1 vs. L2 regularization, and rotational invariance. Proc. ICML https://icml.cc/Conferences/2004/proceedings/papers/354.pdf. Accessed December 15, 2018.

27. Python: A dynamic, open source programming language. Python Software Foundation. Available at https://www.python.org/. Accessed December 15, 2018.

28. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. JMLR 2011;12:2825–30.

29. Brownlee J Machine Learning Mastery With Weka. https://machinelearningmasterv.com/machine-learning-mastery-weka. Accessed December 15, 2018.

30. Le Cessie S, van Houwelingen JC. Ridge Estimators in Logistic Regression. Applied Statistics 1992;41(1): 191–201.

31. Ojha VK, Abraham A, Snášel V. Metaheuristic design of feedforward neural networks: A review of two decades of research. Engineering Applications of Artificial Intelligence 2017;60:97–116.

32. Deep Learning. http://www.scholarpedia.org/article/Deep_Learning. Accessed December 15, 2018.

33. Estimating Continuous Distributions in Bayesian Classifiers. https://arxiv.org/pdf/1302.4964.pdf. Accessed December 15, 2018.

34. Breiman L Random Forests. Machine Learning 2001;45(1):5–32.

35. Blackman DJ, Saraf S, MacCarthy PA, et al. Long-Term Durability of Transcatheter Aortic Valve Prostheses. J Am Coll Cardiol 2019;73(5):537–545. [PubMed: 30732706]

36. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. J Am Coll Cardiol 2017;69:2657–2664. [PubMed: 28545640]

37. Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. Crit Care Med 1998; 26(11):1793–800. [PubMed: 9824069]

38. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart 2018;104(14):1156–1164. [PubMed: 29352006]

PERSPECTIVES

WHAT IS KNOWN?

Transcatheter aortic valve replacement (TAVR) has emerged as a safe and effective therapy for all patients with severe symptomatic aortic stenosis. Current risk prediction tools for in-hospital mortality in patients undergoing TAVR are limited.

WHAT IS NEW?

The National Inpatient Sample (NIS) TAVR score was developed and internally validated using a machine learning approach. The NIS TAVR score outperformed current available TAVR-related outcomes scores. Our study highlights the potential of artificial intelligence in generating robust models to predict adverse outcomes in TAVR patients.

WHAT IS NEXT?

Further validation and application of ML into the day-to-day clinical practice is required to better understand its true value in patients with severe aortic stenosis.
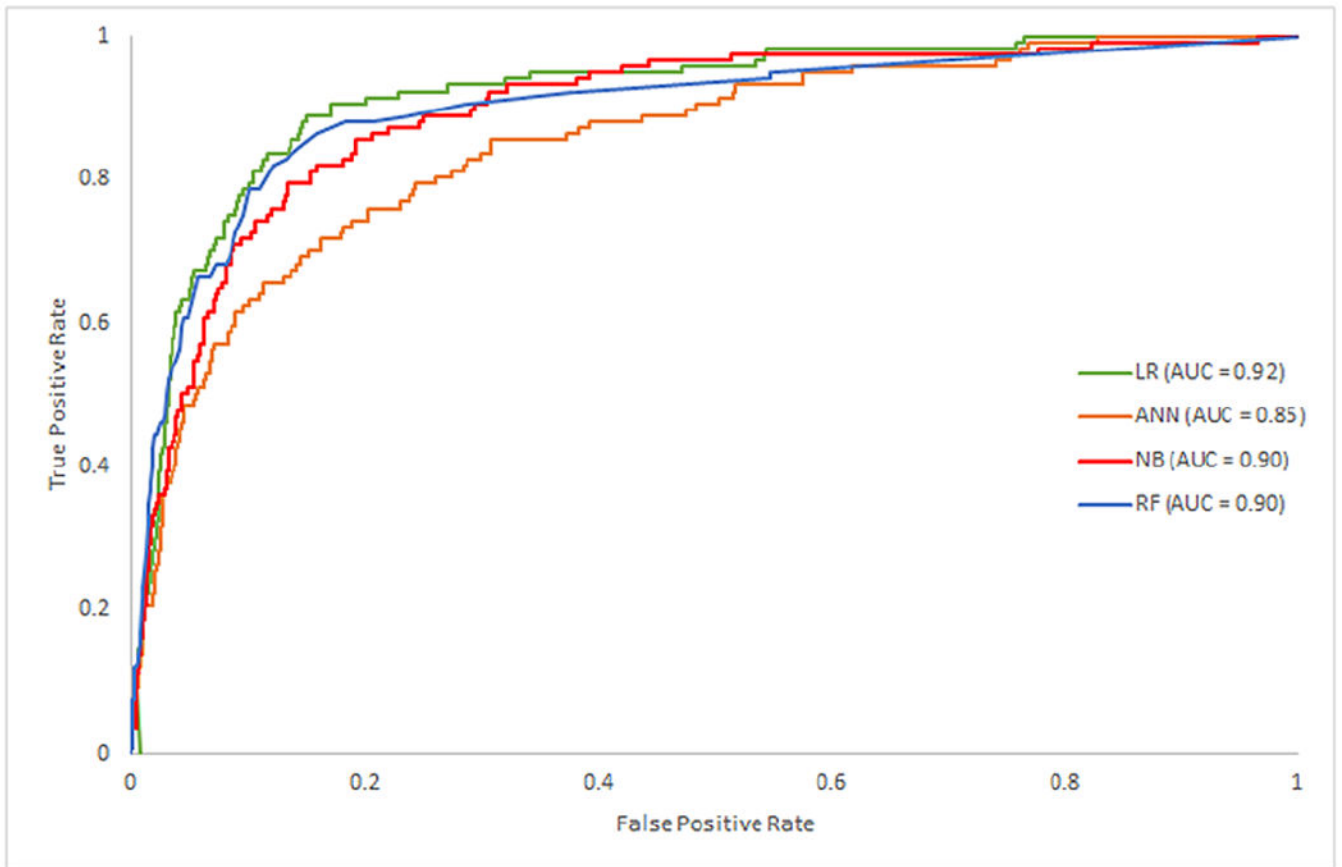
Figure 1.
Operating characteristic area under the curve (AUC) by machine learning model.LR, Logistic Regression; ANN, Artificial Neural Network; NB, Naïve Bayes; RF, Random Forest.
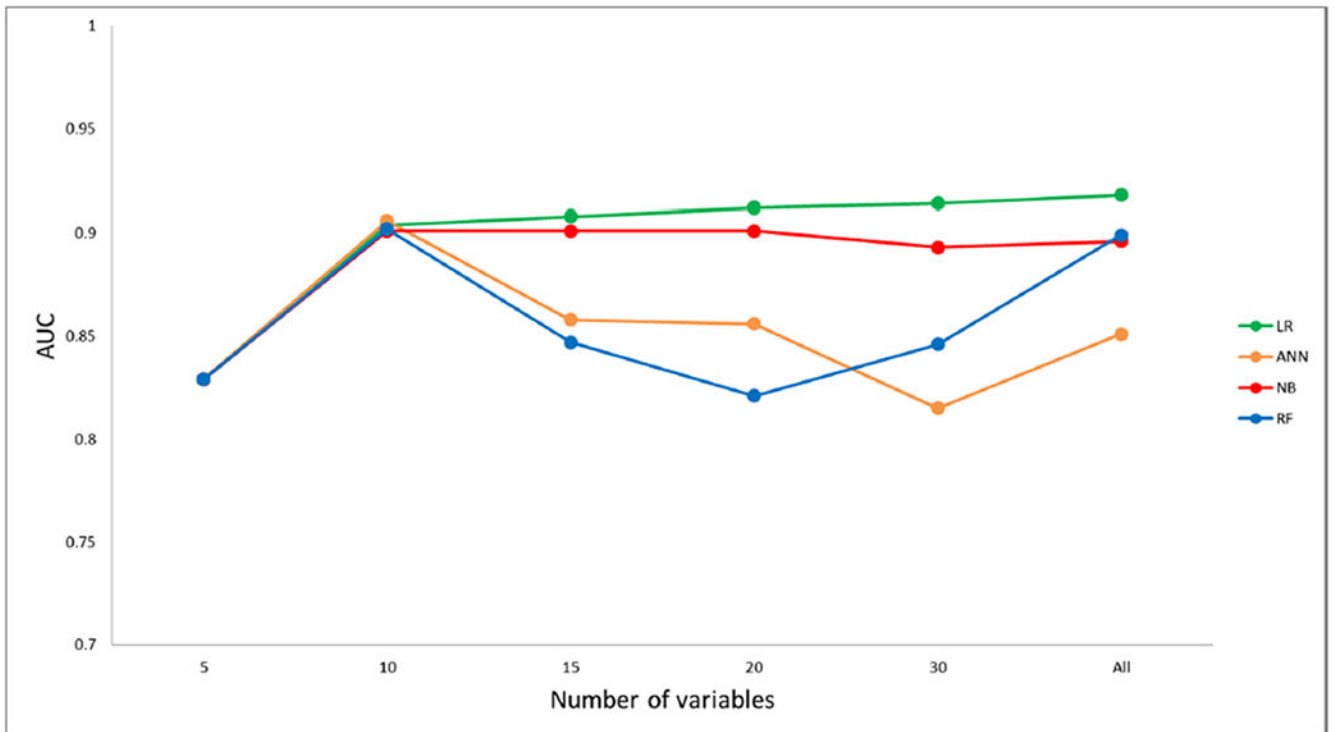
Figure 2.
Predictive performance of TAVR in-hospital mortality models by number of variables.
LR, Logistic Regression; ANN, Artificial Neural Network; NB, Naïve Bayes; RF, Random Forest.

Figure 3.
Machine learning model-based feature importance according to the mean ranking. * LR, Logistic Regression; ANN, Artificial Neural Network; NB, Naïve Bayes; RF, Random Forest.

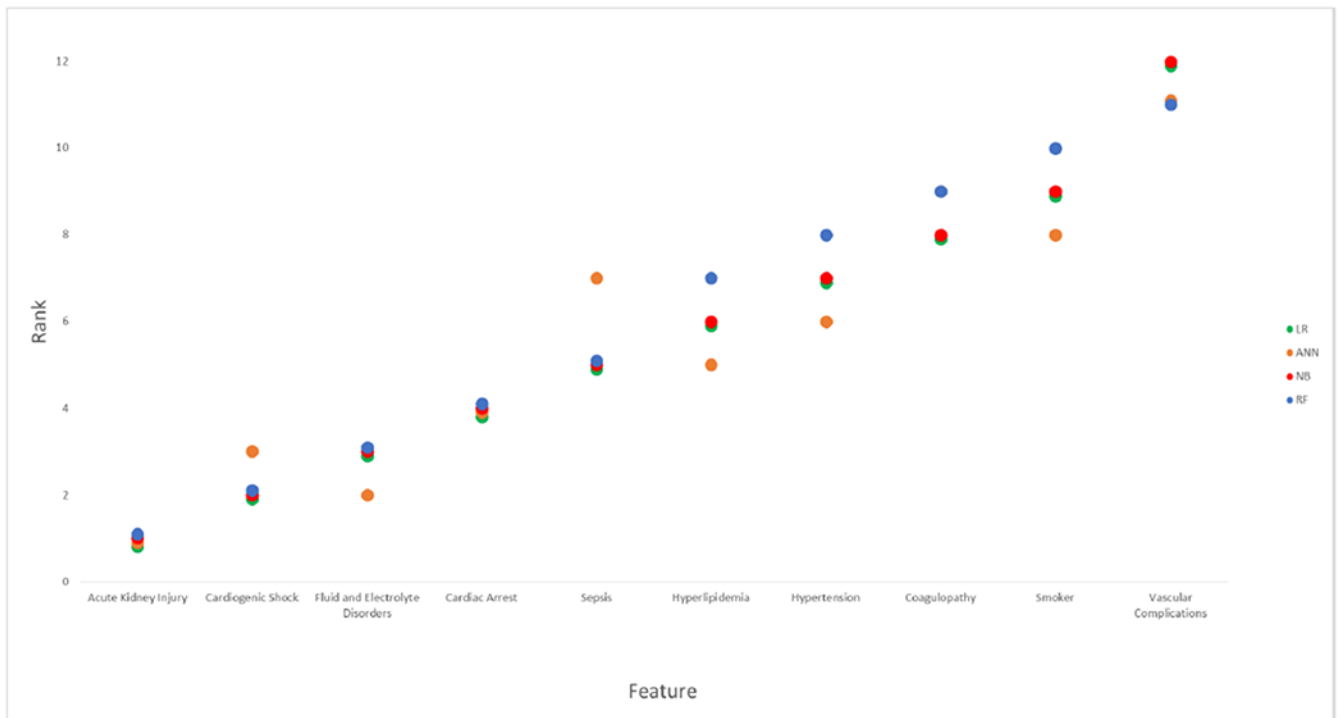*The variable's importance is inversely proportional to the rank. The X-axis depicts the top 10 variables ordered according to mean ranking.
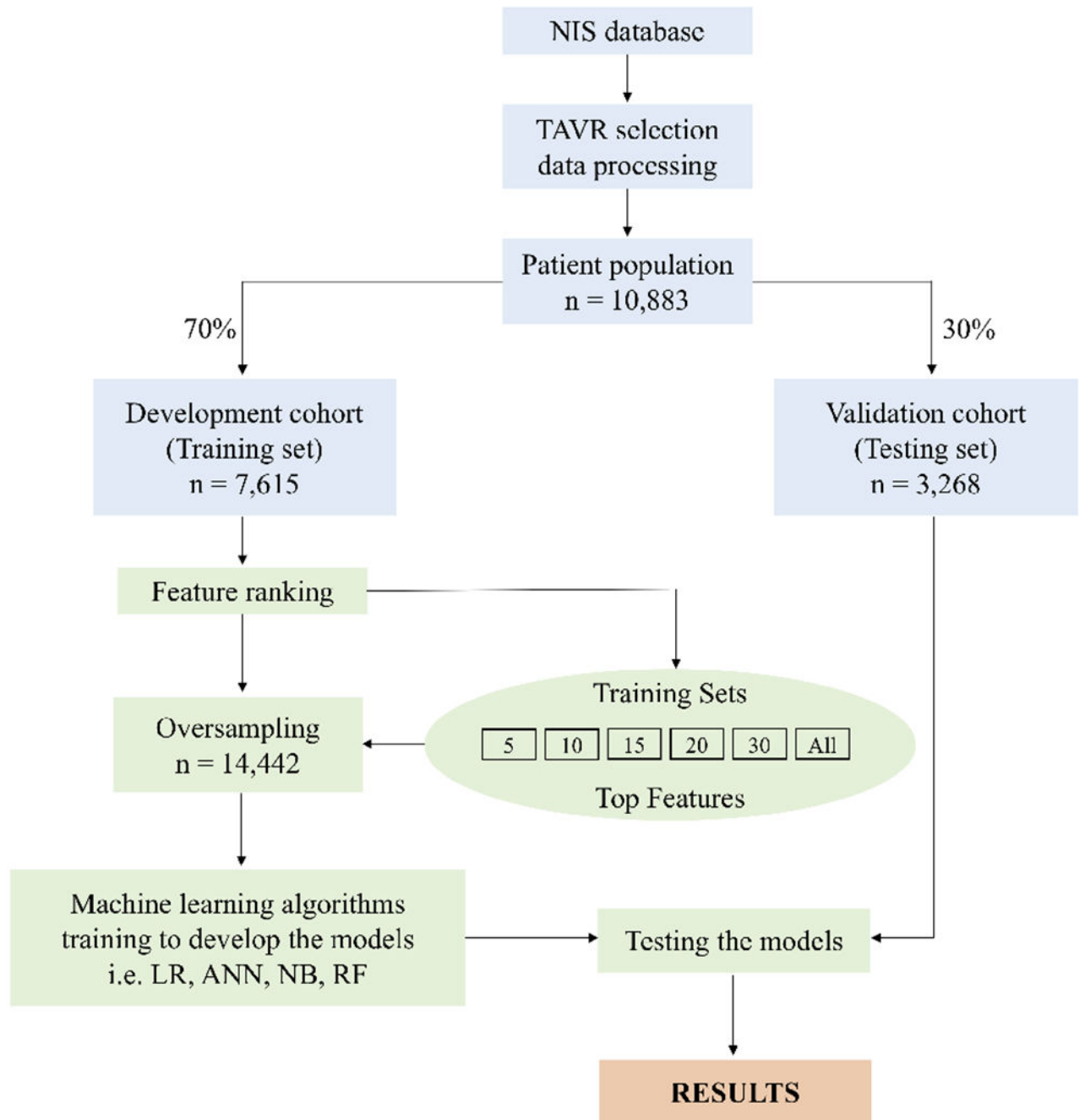
Central Illustration.

Overview of the methods used for data extraction, training and testing. Data from the NIS database was extracted using a tailored scripting in Python to select patients with TAVR. Then the cohort was split into a training set (70% of the data, n=7,615) and test set (30% of the data, n=3,268). A feature ranking method was applied to the training set to determine the top 5,10,15,20,30 variables. Because of data imbalance the training sets were randomly oversampled and then the ML algorithms (i.e. LR, ANN, NB, RF) were trained to develop the models. The different set of variables (including All variables = 43) were used

independently to train each of the ML algorithms. The developed models were validated using the test sets and computing the precision metric results focused on the AUC. NIS, National Inpatient Sample; TAVR, Transcatheter Aortic Valve Replacement; ML, Machine Learning; ANN, Artificial Neural Network; LR, Logistic Regression; NB, Naïve Bayes; RF, Random Forest; AUC, area under the curve..

Table 1.

Demographic and hospital characteristics of TAVR patients.

Variables	Overall n = 10,883	Development cohort n = 7,615	Validation cohort n = 3,268	P-value	Alive n = 10,493	Deceased n = 390	P-value
Demographic characteristics							
Age	81.0±8.5	80.9±8.6	81.2±8.4	.316	80.9±8.5	82.2±7.9	.002
Female	5,191 (47.7)	3,649 (47.9)	1,542 (47.2)	.496	4,972 (47.4)	219 (56.2)	<.001
Smoker	3,250 (29.9)	2,304 (30.3)	946 (29.0)	.179	3,184 (30.3)	66 (16.9)	<.001
Dyslipidemia	7,142 (65.6)	5,012 (65.8)	2,130 (65.2)	.534	6,952 (66.3)	190 (48.7)	<.001
Known CAD	7,798 (71.7)	5,472 (71.9)	2,326 (71.2)	.483	7,554 (72.0)	244 (62.6)	<.001
Prior MI	1,411 (13.0)	993 (13.0)	418 (12.8)	.746	1,370 (13.1)	41 (10.5)	.164
Prior PCI	2,110 (19.4)	1,495 (19.6)	615 (18.8)	.338	2,069 (19.7)	41 (10.5)	<.001
Prior CABG	2,354 (21.6)	1,640 (21.5)	714 (21.9)	.736	2,304 (22.0)	50 (12.8)	<.001
Prior TIA/stroke	1,455 (13.4)	1,025 (13.5)	430 (13.2)	.694	1,416 (13.5)	39 (10.0)	.055
Atrial Fibrillation	4,837 (44.5)	3,376 (44.3)	1,461 (44.7)	.736	4,652 (44.3)	185 (47.4)	.247
Carotid Artery Disease	815 (7.5)	547 (7.2)	268 (8.2)	.070	803 (7.7)	12 (3.1)	.001
Prior PPM	1,141 (10.5)	799 (10.5)	342 (10.5)	.993	1,112 (10.6)	29 (7.4)	.055
Prior ICD	353 (3.2)	246 (3.2)	107 (3.3)	.953	346 (3.3)	7 (1.8)	.134
Elixhauser comorbidities							
Heart Failure	921 (8.5)	640 (8.4)	281 (8.6)	.767	862 (8.2)	59 (15.1)	<.001
Diabetes Mellitus	3,830 (35.2)	2,700 (35.5)	1,130 (34.6)	.391	3,721 (35.5)	109 (28.0)	.003
Hypertension	8,759 (80.5)	6,110 (80.2)	2,649 (81.1)	.334	8,509 (81.1)	250 (64.1)	<.001
Obesity	1,594 (14.7)	1,133 (14.9)	461 (14.1)	.310	1,555 (14.8)	39 (10.0)	.010
Peripheral vascular disease	3,154 (29.0)	2,209 (29.0)	945 (28.9)	.941	3,028 (28.9)	126 (32.3)	.156
Chronic kidney disease	3,812 (35.0)	2,657 (34.9)	1,155 (35.3)	.667	3,637 (34.7)	175 (44.9)	<.001
Anemia	142 (1.3)	92 (1.2)	50 (1.5)	.206	139 (1.3)	3 (0.8)	.470
Chronic Pulmonary Disease	3,598 (33.1)	2,534 (33.3)	1,064 (32.6)	.479	3,471 (33.1)	127 (32.6)	.875
Coagulopathy	2,430 (22.3)	1,719 (22.6)	711 (21.8)	.361	2,286 (21.8)	144 (36.9)	<.001
Liver disease	279 (2.6)	200 (2.6)	79 (2.4)	.571	268 (2.6)	11 (2.8)	.870
Cancer	399 (3.7)	263 (3.5)	136 (4.2)	.081	390 (3.7)	9 (2.3)	.188
Valvular disease	267 (2.5)	182 (2.4)	85 (2.6)	.559	249 (2.4)	18 (4.6)	.008
Pulmonary circulation disorder	292 (2.7)	201 (2.6)	91 (2.8)	.716	264 (2.5)	28 (7.2)	<.001
Fluid and electrolyte disorder	2,683 (24.7)	1,832 (24.1)	851 (26.0)	.030	2,487 (23.7)	196 (50.3)	<.001
Median household income (percentiles)							
0 to 25th	2,323 (21.7)	1,641 (21.9)	682 (21.2)	.942	2,240 (21.7)	83 (21.6)	.648
26th to 50th	2,675 (25.0)	1,869 (24.9)	806 (25.1)		2,583 (25.0)	92 (23.9)	
51th to 75th	2,723 (25.4)	1,904 (25.4)	819 (25.5)		2,632 (25.5)	91 (23.6)	
76th to 100th	3,000 (28.0)	2,093 (27.9)	907 (28.2)		2,881 (27.9)	119 (30.9)	
Hospital characteristics							

Variables	Overall n = 10,883	Development cohort n = 7,615	Validation cohort n = 3,268	P-value	Alive n = 10,493	Deceased n = 390	P-value
Demographic characteristics							
Bed size							
Small	517 (4.8)	345 (4.5)	172 (5.3)	.257	498 (4.8)	19 (4.9)	.991
Medium	2,054 (18.9)	1,439 (18.9)	615 (18.8)		1,980 (18.9)	74 (19.0)	
Large	8,312 (76.4)	5,831 (76.6)	2,481 (75.9)		8,015 (76.4)	297 (76.2)	
Hospital location							
Rural	89 (0.8)	62 (0.8)	27 (0.8)	.970	85 (0.8)	4 (1.0)	.369
Urban non-teaching	1,071 (9.8)	746 (9.8)	325 (9.9)		1,025 (9.8)	46 (11.8)	
Urban teaching	9,723 (89.3)	6,807 (89.4)	2,916 (89.2)		9,383 (89.4)	340 (87.2)	
Hospital region							
Northeast	2,817 (25.9)	1,991 (26.2)	826 (25.3)	.615	2,725 (26.0)	92 (23.6)	.004
Midwest	2,108 (19.4)	1,486 (19.5)	622 (19.0)		2,040 (19.4)	68 (17.4)	
South	3,919 (36.0)	2,727 (35.8)	1,192 (36.5)		3,746 (35.7)	173 (44.4)	
West	2,039 (18.7)	1,411 (18.5)	628 (19.2)		1,982 (18.9)	57 (14.6)	
Transapical TAVR	1,692 (15.6)	1,182 (15.5)	510 (15.6)	.935	1,603 (15.3)	89 (22.8)	<.001
Endovascular TAVR	9,209 (84.6)	6,449 (84.7)	2,760 (84.5)	.780	8,907 (84.9)	302 (77.4)	<.001
Hospital complications							
Stroke/TIA	322 (3.0)	234 (3.1)	88 (2.7)	.312	286 (2.7)	36 (9.2)	<.001
AMI	251 (2.3)	177 (2.3)	74 (2.3)	.903	215 (2.1)	36 (9.2)	<.001
PPM placement	851 (7.8)	602 (7.9)	249 (7.6)	.638	831 (7.9)	20 (5.1)	.055
Conversion to SAVR	18 (0.2)	12 (0.2)	6 (0.2)	.961	14 (0.1)	4 (1.0)	.003
Cardiogenic Shock	411 (3.8)	289 (3.8)	122 (3.7)	.920	282 (2.7)	129 (33.1)	<.001
Cardiac arrest	352 (3.2)	248 (3.3)	104 (3.2)	.887	247 (2.4)	105 (26.9)	<.001
Major bleeding	287 (2.6)	199 (2.6)	88 (2.7)	.863	252 (2.4)	35 (9.0)	<.001
Vascular complications	299 (2.8)	217 (2.9)	82 (2.5)	.351	251 (2.4)	48 (12.3)	<.001
AKI	1,893 (17.4)	1,323 (17.4)	570 (17.4)	.953	1,672 (15.9)	221 (56.7)	<.001
Sepsis	238 (2.2)	164 (2.2)	74 (2.3)	.771	151 (1.4)	87 (22.3)	<.001

Data are presented as Mean ± SD or n (%). Abbreviations: CAD, Coronary Artery Disease; MI, Myocardial Infarction; PCI, Percutaneous Coronary Intervention; CABG, Coronary Artery Bypass Graft; TIA, Transient Ischemic Attack; PPM, Permanent Pacemaker; ICD, Implantable Cardioverter Defibrillator; SAVR, surgical aortic valve replacement; AKI, Acute Kidney Injury.

Table 2.

Classifiers' predictive performance in the validation cohort (testing set).

Classifier	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F-Score
Logistic Regression	87.7	83.9	96.5	83.8	0.92
Artificial Neural Networks	94.4	42.2	95.3	40.2	0.95
Naïve Bayes	84.8	79.7	96.2	79.5	0.90
Random Forest	96.3	16.7	95.0	13.7	0.96

PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Sensitivity: Ability of the trained models (i.e. LR, ANN, NB, RF) to correctly identify patients WITH the event (i.e. 'deceased'). *Specificity:* Ability of the trained models (i.e. LR, ANN, NB, RF) to correctly identify patients WITHOUT the event (i.e. 'alive'). *PPV:* Proportion of patients identified WITH the event (i.e. 'deceased') that truly had the event. *NPV:* Proportion of patients identified WITHOUT the event (i.e. 'alive') that truly did not have the event.

Table 3.

Characterization of currently available TAVR mortality risk scores.

Score(year)	Population	Primary outcome	AUC
FRANCE-2 (2014)	Total: n=3,833 Derivation cohort: n=2,552 Validation cohort: n=1,281	30-day mortality	0.59
TARIS (2014)	Total: n=1,178 Derivation cohort: n=845 Validation cohort: n=333	30-day mortality	0.69
OBSERVANT (2014)	Total: n=1,878 Derivation cohort: n=1,256 Validation cohort: n=622	30-day mortality	0.71
TAVI$_2$ (2015)	Total: n=511 Derivation cohort: n=511 Validation cohort: n=100 [†]	1-year mortality	0.72
CoreValve US (2016)	Total: n=3,687 Derivation cohort: n=2,482 Validation cohort: n=1,205	30-day and 1-year mortality	0.75 and 0.79, respectively
STS/ACC TVT (2016)	Total: n=20,540 Derivation cohort: n=13,672 Validation cohort: n=6,868	In-hospital mortality	0.66
UK TAVI (2017)	Total: n= 6,339 Derivation cohort: n=6,339 Validation cohort: n=2,969 [†]	30-day mortality	0.66
CAPRI (2019)	Total: n= 1,736 Derivation cohort: n=1,425 Validation cohort: n=311	1 -year mortality	0.68
NIS TAVR (2019) [*]	Total: n= 10,891 Derivation cohort: n=7,624 Validation cohort: n=3,667	In-hospital mortality	0.92

[†] Validation sample was obtained from the derivation cohort.

[*] Best model obtained in our study.