JACC: ADVANCES VOL. 3, NO. 9, 2024

ORIGINAL RESEARCH

VALVULAR HEART DISEASE

Machine Learning Prediction for Prognosis of Patients With Aortic Stenosis

Sara Shimoni, MD,[a] Ruslan Sergienko, MBBS,[b] Pablo Martinez-Legazpi, PhD,[c] Valery Meledin, MD,[a] Sorel Goland, MD,[a] Sagie Tshori, MD,[a] Jacob George, MD,[a] Javier. Bermejo, MD,[d,*] Lior Rokach, PhD[d,*]

ABSTRACT

BACKGROUND Aortic valve stenosis of any degree is associated with poor outcomes.

OBJECTIVES The authors aimed to develop a risk prediction model for aortic stenosis (AS) prognosis using machine learning techniques.

METHODS A prognostic algorithm was developed using an AS registry of 10,407 patients undergoing echocardiography between 2008 and 2020. Clinical, echocardiographic, laboratory, and medication data were used to train and test a time-to-event model, the random survival forest (RSF), for AS patient's prognosis. The composite outcome included aortic valve replacement or mortality. The SHapley Additive exPlanations method attributed the importance of variables and provided personalized risk assessment. The algorithm was validated in 2 external cohorts of 11,738 and 954 patients with AS.

RESULTS The median follow-up of the primary cohort was 48 (21-87) months. In this period, 1,116 patients underwent aortic valve replacement, and 5,069 patients died. RSF had an area under the curve (AUC) of 0.83 (95% CI: 0.80-0.86) and 0.83 (95% CI: 0.81-0.84) for outcomes prediction at 1 and 5 years, respectively. Using a cut-off of 50%, the RSF sensitivity and specificity for the composite outcome, were 0.80 and 0.73, respectively. Validation performance in the 2 external cohorts was similar, with AUCs of 0.73 (95% CI: 0.72-0.74) and 0.74 (95% CI: 0.72-0.76), respectively. AS severity, age, serum albumin, pulmonary artery pressure, and chronic kidney disease emerged as the top significant variables in the model.

CONCLUSIONS In patients with AS, a machine learning algorithm predicts outcomes with good accuracy, and prognostic characteristics were identified. The model can potentially guide risk factor modification and clinical decisions to improve patient prognosis. (JACC Adv. 2024;3:101135) © 2024 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

From [a]The Heart Institute, Kaplan Medical Center, Rehovot, Israel and Hebrew University and Hadassah Medical School, Jerusalem, Israel; [b]Department of Health Policy and Management, Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel; [c]Department of Mathematical Physics and Fluids, Facultad de Ciencias, Universidad Nacional de Educación a Distancia, UNED, and CIBERCV, Madrid, Spain; and the [d]Department of Cardiology, Hospital General Universitario Gregorio Marañón, Facultad de Medicina, Universidad Complutense de Madrid, Instituto de Investigación Sanitaria Gregorio Marañón, and CIBERCV, Madrid, Spain *Drs Bermejo and Rokach contributed equally to this work.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received February 27, 2024; revised manuscript received May 28, 2024, accepted June 7, 2024.

ISSN 2772-963X https://doi.org/10.1016/j.jacadv.2024.101135

ABBREVIATIONS AND ACRONYMS

AS = aortic stenosis

AVA = aortic valve area

AVR = aortic valve replacement

AUC = area under the curve

EMR = electronic medical record

LV = left ventricle

ML = machine learning

PAP = pulmonary artery systolic pressure

RSF = random survival forest

ST = survival trees

The prevalence of aortic stenosis (AS) is increasing due to the aging population.[1] Although severe AS can be fatal without aortic valve replacement (AVR), surgical and transcatheter valve replacement significantly improve prognosis.[2] However, AVR is indicated in patients with severe symptomatic AS or severe AS with reduced left ventricular systolic function. Numerous studies have shown that even mild AS is associated with poor clinical outcomes.[3,4] The links between AS of any degree and mortality may be, to some extent, related to comorbidities. However, when adjusted for baseline confounders, the risk observed at mild degrees of valvular obstruction remains significant.

Recently, machine learning (ML) methods have been adopted in cardiovascular research community, and they have been applied successfully in valvular disease as well, to predict survival or grade the disease.[5,6] Most of the ML studies addressing patients with AS are based on relatively simple classification and clustering approaches. Also, these studies are focused mainly on patients with moderate or severe AS and include only routine echocardiographic measurements. Since any degree of AS is related to impaired outcome, there is a need to develop models to predict risks and outcomes in patients with various degrees of AS that include the comorbidities. Our objective was to develop a risk prediction model for AS within a large registry of patients with wide range of AS severities, combining ultrasound and electronic medical record (EMR) information, using advanced ML techniques. Model validation was performed in 2 independent cohorts.

METHODS

PATIENTS AND CLINICAL DATA. The algorithm was developed using data from the Kaplan Medical Center AS Patients Prospective Registry. This registry comprises patients with AS who underwent echocardiography between 2008 and 2020. We included individuals with at least one echocardiogram with sufficient data on AS severity, in accordance with guidelines.[7,8] Patients with AVR before the baseline echocardiography study were excluded. Initially, the patient cohort consisted of 12,757 patients. We excluded 1,259 patients due to previous AVR, 121 patients due to subvalvular gradients, and 970 patients due to incomplete echocardiographic data (Supplemental Figure 1). The final cohort comprised 10,407 patients: 7,173 with mild AS, 2,007 with moderate AS, and 1,227 with severe AS. The study protocol was approved by the Kaplan Medical Center Review Board and Ethics Committee, which waived the need for informed consent.

Mild AS was defined by an aortic valve area (AVA) of 1.5 to 2.0 cm^2 and a peak velocity of 2.5 to 2.9 m/s. Moderate AS was characterized by an AVA of 1.0 to 1.5 cm^2 and a peak velocity of 3 to 4 m/s, while severe AS was defined as an AVA <1.0 cm^2 and a peak velocity >4 m/s. In cases of discordant AS severity, we re-evaluated measurements and measured calcium scores.[9] The aortic valve calcium score was measured only in patients with discordant AS severity and was not included in the model. Left and right chamber dimensions and function, valvular assessment, and pulmonary artery pressure (PAP) measurements were performed according to established guidelines.[10,11] Some of the classic diastolic function indices were excluded from the analysis due to a high percentage of patients with atrial fibrillation and severe mitral annular calcifications.[12] However, we included left atrial size and PAP as surrogates for diastolic function.

Clinical data included age, gender, relevant implantable cardioverter-defibrillator-10 codes for cardiovascular diseases, and comorbidities such as chronic lung disease, dementia, and osteoporosis, obtained from the EMRs. We collected data on all-cause mortality and AVR. Laboratory tests were available in 80% of the patients, and all medications were included in the dataset. Baseline data accounted for the first echocardiography and its respective clinical data, collected in a 3-month window. The primary outcome of the model was a combined end point of AVR and all-cause mortality. Patients received follow-up and treatment from their primary physicians and cardiologists, and any intervention was performed in accordance with guidelines and clinical judgment.

VALIDATION COHORTS. The model was validated twice using: first, a cohort of 11,738 patients with AS obtained from the Clalit Health Services database (Clalit HMO). These patients underwent echocardiography between 2013 and 2022. Clalit Health Services provides care for 4.7 million patients, representing 52% of the total Israeli population. The clinical data, including demographics, risk factors, and comorbidities were recorded as in the original data from the Kaplan Medical Center. However, echocardiography reports were available only for a random selection of patients, and the electronic reports had limited data. Specifically, information on aortic and tricuspid regurgitations was not available,

nor was data on PAP. The second cohort consisted of 954 patients with AS from the Hospital General Universitario Gregorio Marañón (HGUGM) in Madrid, Spain. These patients underwent echocardiography between July 1, 2015, and June 30, 2016. Clinical and echocardiographic data for this cohort have already been reported.[4]

MACHINE LEARNING MODEL. While most of the previous studies that used ML techniques focused on classification tasks, we addressed the task as time-to-event. This enables us to handle censored data by incorporating time-dependent impact of covariates and enabling a direct estimation of survival probabilities. All clinical and echocardiographic data were used to train a time-to-event model for the prognosis of AS patients using a random survival forest (RSF).[13,14] RSF is a powerful ensemble method that combines the principles of random forests with survival analysis techniques. The RSF model is constructed by assembling a collection of survival trees (ST). Survival trees are specifically designed to handle censored data, which is common in survival analysis where some events of interest have not yet occurred or have been lost to follow-up. The structure of ST is built by recursively partitioning the data based on the independent variables. Each ST is built using a random subset of patients. At each node of the tree, the algorithm identifies the most informative variable (eg, age) and its corresponding threshold (eg, 80) and then separates the patients into branches (eg, age <80 vs age ≥80 years) such that patients in each branch have similar survival patterns. Each tree predicts a survival function—the probability of surviving up to a certain time "*t*." To make an overall prediction for new patients, their data will run through each tree to obtain multiple survival function estimates. These are averaged across trees to produce the model's consensus prediction. An illustrative shallow tree is shown in Supplemental Figure 2. The details of ML method and statistical analysis are summarized in Supplemental Methods.

RESULTS

A total of 10,407 patients, aged 77 (69-84) years, with 52% being women, were included in the primary cohort. Sixty-nine percent of the patients had mild AS, while 19% had moderate AS and 12% had severe AS. The baseline characteristics of the patients are shown in **Table 1** and Supplemental Table 1. Patients showed frequent comorbidities. Left ventricular (LV) size and left ventricular ejection fraction were preserved, but there was evidence of left ventricular hypertrophy, a dilated left atrium, and elevated PAP.

FOLLOW-UP. The median follow-up of the primary cohort was 48 (21-87) months. During this period, 1,116 patients underwent AVR, and 5,069 patients died, 463 of them after AVR. The Kaplan-Meier analysis of the outcomes according to baseline AS severity grade is shown in Supplemental Figures 3A to 3C. The clinical and echocardiographic predictors for the combined end point of AVR and mortality are shown in Supplemental Table 2. Many clinical and echocardiographic parameters were related to outcome. Among others, moderate (hazard ratio: 1.57; 95% CI: 1.38-1.52; $P < 0.001$) and severe (hazard ratio: 2.50; 95% CI: 2.32-2.70; $P < 0.001$) AS were 2 of the most relevant predictors of outcome.

DEVELOPING A MACHINE LEARNING MODEL FOR PREDICTING OUTCOMES IN PATIENTS WITH AORTIC STENOSIS. The RSF model showed an area under the curve (AUC) of 0.83 (95% CI: 0.80-0.86) and 0.83 (95% CI: 0.81-0.84) to predict the composite outcome at 1 and 5 years, respectively (**Central Illustration**). The AUC of RSF for prediction outcome at 1 year was significantly higher than for the conventional Cox proportional-hazard model (0.77 [95% CI: 0.76-0.79], $P = 0.02$) (**Figure 1A**). Using a default cut-off of 50% (above and below median), the RSF sensitivity and specificity for end point predictions were 0.80 and 0.73, respectively, at the end of the first year. The RSF model was also superior to Cox proportional-hazard model for predicting separate outcomes of all-cause mortality (AUC: 0.77 [95% CI: 0.74-0.8] vs 0.73 [95% CI: 0.71-0.75], respectively, $P = 0.014$, **Figure 1B**) and AVR (AUC: 0.92 [95% CI: 0.90-0.94] vs 0.89 [95% CI: 0.86-0.92], respectively, $P = 0.033$) (**Figure 1C**).

EXTERNAL VALIDATION OF THE MODEL. Characteristics of both validation cohorts are shown in **Table 1** and Supplemental Table 1. The Clalit HMO database comprised 11,738 AS patients with a median age of 78 (IQR: 64-89) years, and 54% were male. The median follow-up was 39 (IQR: 16-62) months. During this period, 2,419 patients underwent AVR, and 4,729 patients died. Significant differences exist between the patient profiles in this database compared to the Kaplan Medical Center database. The proportion of males was higher, and there were differences in terms of risk factors, comorbidity frequencies, and the prevalence of moderate and severe AS cases within this database. Nevertheless, despite these differences and the absence of certain echocardiographic parameters, ML RSF model still demonstrated a high level of accuracy, exhibiting a significant improvement in predictive accuracy when compared to the Cox proportional-hazards model within this database (AUC at 1-year: RSF 0.73 [95% CI: 0.72-0.74], vs COX

TABLE 1 Main Data of the Training and Validation Cohorts

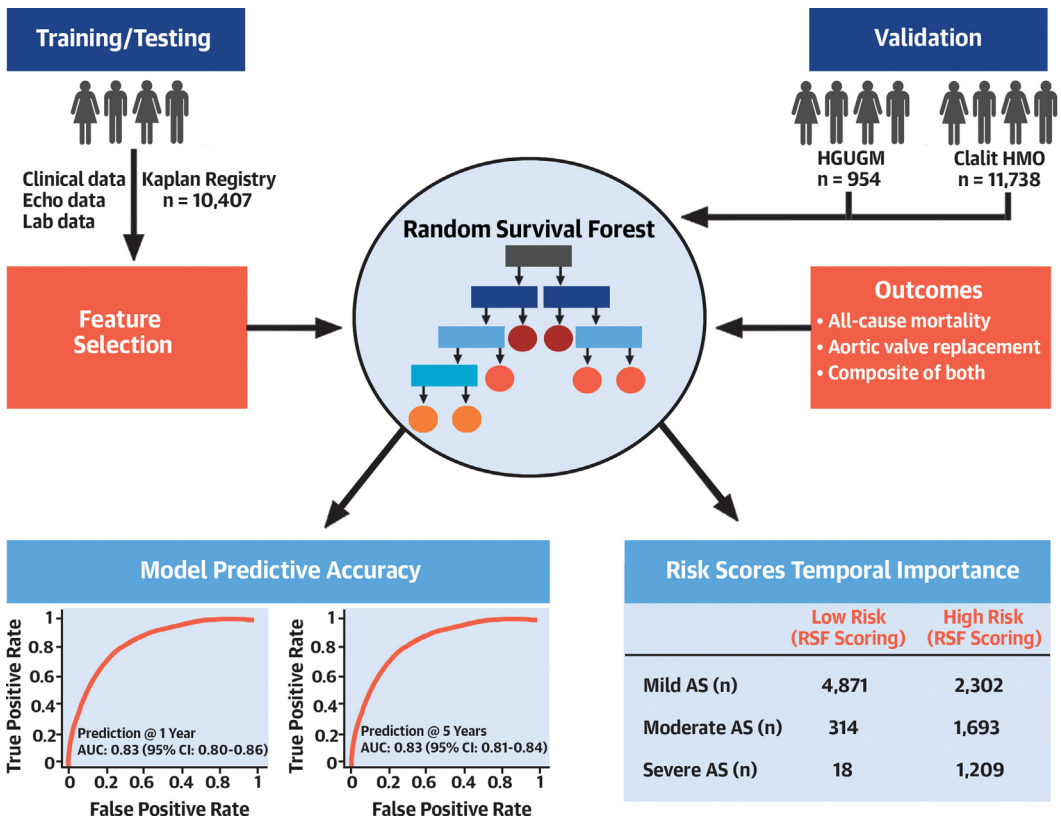
	Kaplan Hospital (n = 10,407)	Clalit HMO (n = 11,738)	HGUGM (n = 954)
Demographics			
Age (y)	77 (69-84)	78 (69-84)	78 (64-89)
Male	4,957 (47.7)	6,317 (53.8)	444 (46.60)
Hypertension	8,539 (82.1)	9,878 (84.2)	620 (64.7)
Diabetes mellitus	3,442 (33.1)	4,615 (39.3)	161 (16.7)
Hyperlipidemia	7,236 (69.6)	9,446 (80.5)	
Heart failure	3,060 (29.4)	4,931 (42.0)	371 (38.8)
History of CABG or PCI	1,945 (19)	3,211 (27)	287 (28.8)
History of acute coronary syndrome	353 (3.4)	768 (6.5)	87 (9.1)
Atrial fibrillation	2,713 (26.1)	3,893 (33.2)	
PVD/CVA	661 (6.4)	918 (7.8)	289 (28.5)
Mitral or tricuspid valve surgery	16 (0.2)	51 (0.4)	
Pacemaker procedure	450 (4.3)	623 (5.3)	
History of myocarditis	86 (0.8)	218 (1.9)	
Chronic kidney disease	1,462 (14.1)	2,543 (21.7)	366 (38.3)
Pulmonary embolism/DVT	165 (1.6)	298 (2.5)	
Dementia/Alzheimer's	703 (6.8)	898 (7.7)	21 (2.2)
Echocardiography			
Peak aortic valve gradient (mm Hg)	30 ± 16	45 ± 22	37 ± 24
Mean aortic valve gradient (mm Hg)	19 ± 19	28 ± 15	24 ± 15.3
Aortic valve area (cm^2)	1.46 ± 0.39	1.00 ± 0.34	1.37 ± 0.6
LVED diameter (mm)	46.08 ± 6.28	46.59 ± 6.48	45.95 ± 7.91
LVES diameter (mm)	28.65 ± 6.80	30.53 ± 7.29	31.80 ± 7.23
Posterior wall thickness (mm)	11.26 ± 1.88	10.56 ± 1.66	10.70 ± 2.20
Interventricular septum thickness (mm)	12.42 ± 2.09	11.76 ± 2.08	
Left atrial diameter (mm)	40.22 ± 7.00		41.33 ± 8.26
Left atrial area (cm^2)	22.57 ± 5.50	22.32 ± 6.56	
Aortic root diameter (mm)	30.40 ± 30.40	28.52 ± 28.52	
Ascending aorta diameter (mm)	33.53 ± 4.59	35.35 ± 5.17	
LVEF (%)	53.64 ± 8.51	55.41 ± 11.14	56.06 ± 9.59
Pulmonary artery systolic pressure (mm Hg)	40.5 ± 13.6		35.8 ± 15.2
Mitral valve regurgitation			
No/trace	1,751 (16.8)		4,34 (45.5)
Mild/mild to moderate	7,678 (73.8)	5,322 (45.3)	267 (28)
Moderate	710 (6.8)	4,575 (39.0)	135 (14)
Moderate to severe/severe	268 (2.6)	1,841 (15.7)	118 (12.4)
Aortic valve regurgitation			
No/trace	4,567 (43.9)		807 (84.6)
Mild/mild to moderate	5,418 (52.1)		72 (7.5)
Moderate	334 (3.2)		21 (2.2)
Moderate to severe/severe	88 (0.8)		54 (5.7)
Tricuspid valve regurgitation			
No/trace	1,529 (14.7)		
Mild/mild to moderate	7,898 (75.9)		
Moderate	728 (7.0)		
Moderate to severe/severe	252 (2.4)		

Values are median (IQR), n (%), or mean ± SD.

CABG = coronary artery bypass grafting; CVA = cerebrovascular accident; DVT = deep vein thrombosis; HGUGM = Hospital General Universitario Gregorio Marañón; LVED = left ventricular end-diastolic; LVEF = left ventricular ejection fraction; LVES = left ventricular end-systolic; PCI = percutaneous coronary intervention; PVD = peripheral vascular disease.

0.69 [95% CI: 0.68-0.70], $P < 0.0001$) (Figure 2A). The HGUGM validation cohort consisted of 954 AS patients with a median age of 78 (IQR: 64-89) years, and 48% were male. The median follow-up duration was 43 (IQR: 18-72) months. Over the follow-up period, 92 patients underwent AVR, and 470 patients died. As seen in Table 1, the patients in this cohort exhibited differences compared to the Kaplan database. There were some variations in comorbidities, and certain variables were missing from this database. The

CENTRAL ILLUSTRATION **Machine Learning Prediction for Prognosis of Patients With Aortic Stenosis**



Shimoni S, et al. JACC Adv. 2024;3(9):101135.

A time-to-event random survival forest model was developed to predict patients' prognosis. The composite outcome included aortic valve replacement (AVR) or mortality. The algorithm was validated in 2 external cohorts of 11,738 and 954 patients with receiver operating characteristic curves for assessing random survival forest (RSF) accuracy to predict 1-year and 5-year outcomes. The model-based patient scores show that high- and low-risk patients have various severities of aortic valve stenosis. AS = aortic stenosis

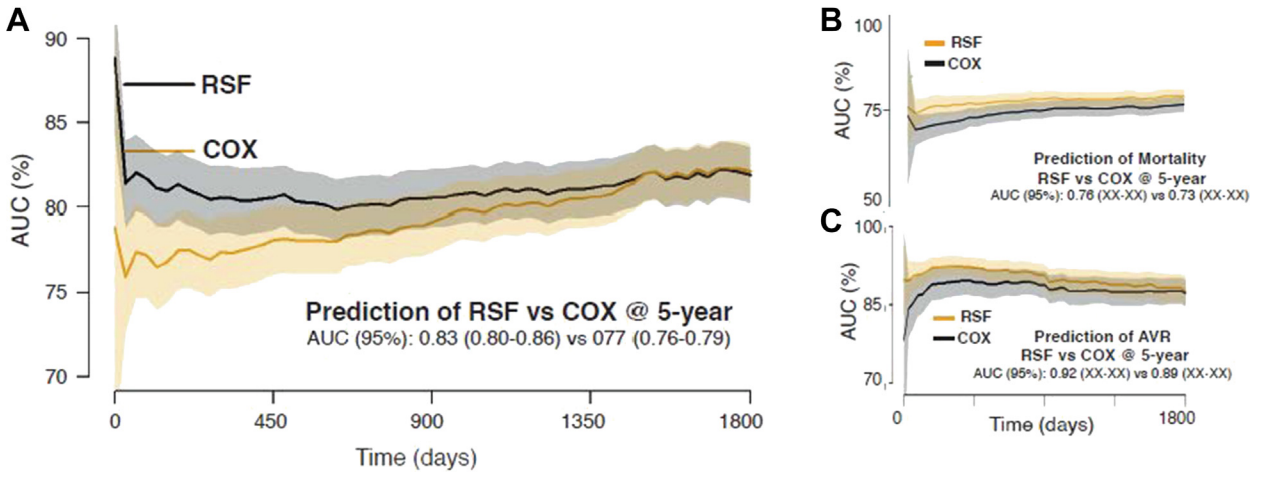
model's accuracy remained robust in this cohort as well (AUC at 1-year: RSF 0.74 [95% CI: 0.72-0.76] vs COX 0.72 [95% CI: 0.68-0.72], $P < 0.002$) (Figure 2B).

OUTCOME RISK PREDICTION. We calculated a relative risk score for each patient in the registry. Figure 3A displays the risk scores for patients with mild, moderate, and severe AS. Notably, although risk score is directly proportional to AS severity, some patients with mild AS were found to have high risk scores, and certain patients with severe AS showed lower risk scores than expected. For the sake of illustration, we divided the patients into 2 (Central Illustration) and 4 groups based on the ML model's quartile risk score. Figures 3B to 3D shows the time evolution of each risk score, adjusted by age and sex, in the 3 studied cohorts (all $P < 0.001$).

VARIABLE TIME-DEPENDENT IMPACT ASSESSMENT. Using the time-dependent Brier score for right-censored data, we estimated the average impact of each variable on the predictive performance of the model (Figure 4A). The variables were scored by importance over time. We noted an incremental effect of age over time, for example, while some effects, such as the effect of mild AS, are decreasing over time, probably due to the progression of the aortic valve severity with time.

VARIABLE SELECTION BY IMPORTANCE. The ranking of variables for the entire prognostic model was determined using the SHapley Additive exPlanations (SHAP) method, as displayed in Supplemental Table 3. AS severity, age, serum albumin, PAP, and chronic kidney disease emerged as the top 5

FIGURE 1 Model Performance



(A) Comparison of predictive performance (AUC over time) of RSF model and Cox proportional-hazards survival model to predict combined outcome of AVR and all-cause mortality over time in the primary cohort. (B and C) Comparison of RSF performance (AUC over time) vs Cox proportional hazards to predict all-cause mortality (B) and AVR (C) over time. AUC = area under the curve; AVR = aortic valve replacement; RSF = random survival forest.

significant variables in the model. The significance of each variable in predicting 1- and 5-year prognosis based on variable values is shown in Figure 4B. AS severity proves to be an important variable to predict both 1- and 5-year prognosis. Age is a significant predictor in the short term and gains greater importance in the long-term prognosis. The most relevant echocardiographic predictors are PAP, LV function, hypertrophy, mitral annular calcification, and left atrial size. Clinical predictors include albumin level, renal failure, anemia, and serum calcium levels. Using the SHAP model, we are able to predict the specific variables related to an individual patient's prognosis and forecast the effect of certain baseline variables over time, as shown in Figure 4C. The diagram illustrates a SHAP analysis to predict outcome of an 80-year-old patient with mild AS. As highlighted in this specific example, age, LV hypertrophy, and borderline serum albumin are negative outcome-related variables over time.

The effect of time of baseline study on the model. Patients were studied from 2008 to 2020. During this time period, there were significant changes in AS patients' management, including the introduction of transcatheter AVR, new medications, etc. To address these changes, we tested the model in 2 different time periods: 1) 2008 to 2013; and 2) 2014 to 2020. Training separate models for each period barely improved the predictive performance of the original model (0.245% increase in weighted AUC at 1 year, $P = 0.351$). Alternatively, we included in a single RSF, the time frame as an additional variable, yielding similar results (0.22% increase in weighted AUC at 1 year, $P = 0.386$). Interestingly, when examining the crude (nonadjusted) Kaplan-Meier curves, we observed a prolonged time-to-event among patients assigned to the late period (2014-2020, $P = 0.00037$). To address this issue, we conducted an additional test by training an adversarial classifier (a random forest classifier) to distinguish between the first and second periods using patients' clinical data. This model differentiates well between the 2 periods (with an AUC of 0.838). This suggests that the period information is already encoded within the existing variables of the entire model, as exemplified by the predominant administration of direct oral anticoagulant and protein convertase subtilisin/kexin type 9 inhibitors, among others, in the second period.

Initial model comprises 81 clinical and echocardiographic variables typically conducted and found in the medical records of elderly patients with AS. To streamline the model, we evaluated its predictive accuracy using only 12 variables using the nested cross-validation method. As seen in Supplemental Figure 4, accuracy was not compromised when using the reduced model. However, it's worth noting that this simplified model is cohort-dependent, and the variables may vary with the dataset.

The extra features (beyond the top 12) do not provide a significant contribution to the model's predictive capability. This suggests that these additional features might be redundant or irrelevant. We

maintained the model with all features despite achieving identical accuracy with a reduced feature set. This approach could offer benefits such as enhanced robustness to future data changes and stability in model performance, especially if one of the top features is missing or noisy. This is important due to differences in various EMRs.

DISCUSSION

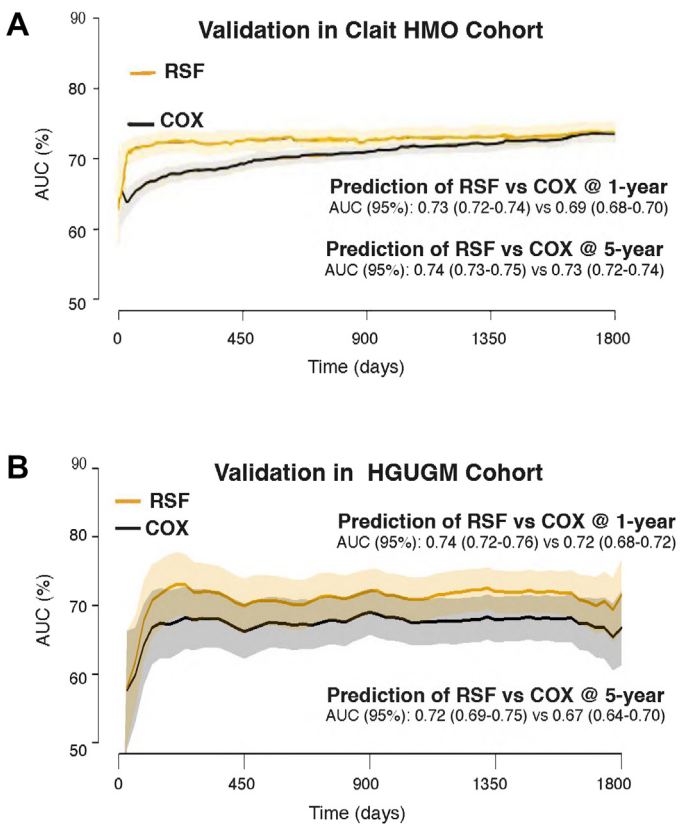
To our knowledge, this is the first study to use an advanced ML framework for generating time-to-event outcome predictions based on clinical data in patients across all grades of AS. The model is built upon a large, single-center dataset containing comprehensive EMR and echocardiographic "real-world" data that was validated on 2 external populations, demonstrating good prognostic accuracy. Furthermore, the model can be employed to assess time-dependent impact and importance of variable, predict individual patient risk, and variables related to patient prognosis. This potential opens the door to personalized, patient-tailored follow-up and treatment.

AS ranks as the third most common cardiovascular disease in the Western world, leading to premature mortality and impacting quality-adjusted life years.[15] Previous studies have demonstrated that AS, regardless of its severity, is linked to premature mortality, particularly among older patients.[4,15] The findings of this study align with these observations, indicating a high percentage of adverse outcome events in patients with all AS grades. Like other valvular diseases, AS is frequently underdiagnosed and undertreated.[16] The relatively low number of AVR and the relatively high number of deaths in our study are consistent with findings in other reports.[3,4]

We used the Cox model to predict patient outcomes in our registry, which included patients with various degrees of AS and a large amount of clinical information. The Cox proportional-hazards model yielded relatively good performance for predicting patient outcomes. However, dataset analyses using Cox proportional-hazards or any other generalized linear models require careful model construction, which is often impractical when dealing with a very large number of predictors and complex unknown interactions.[17]

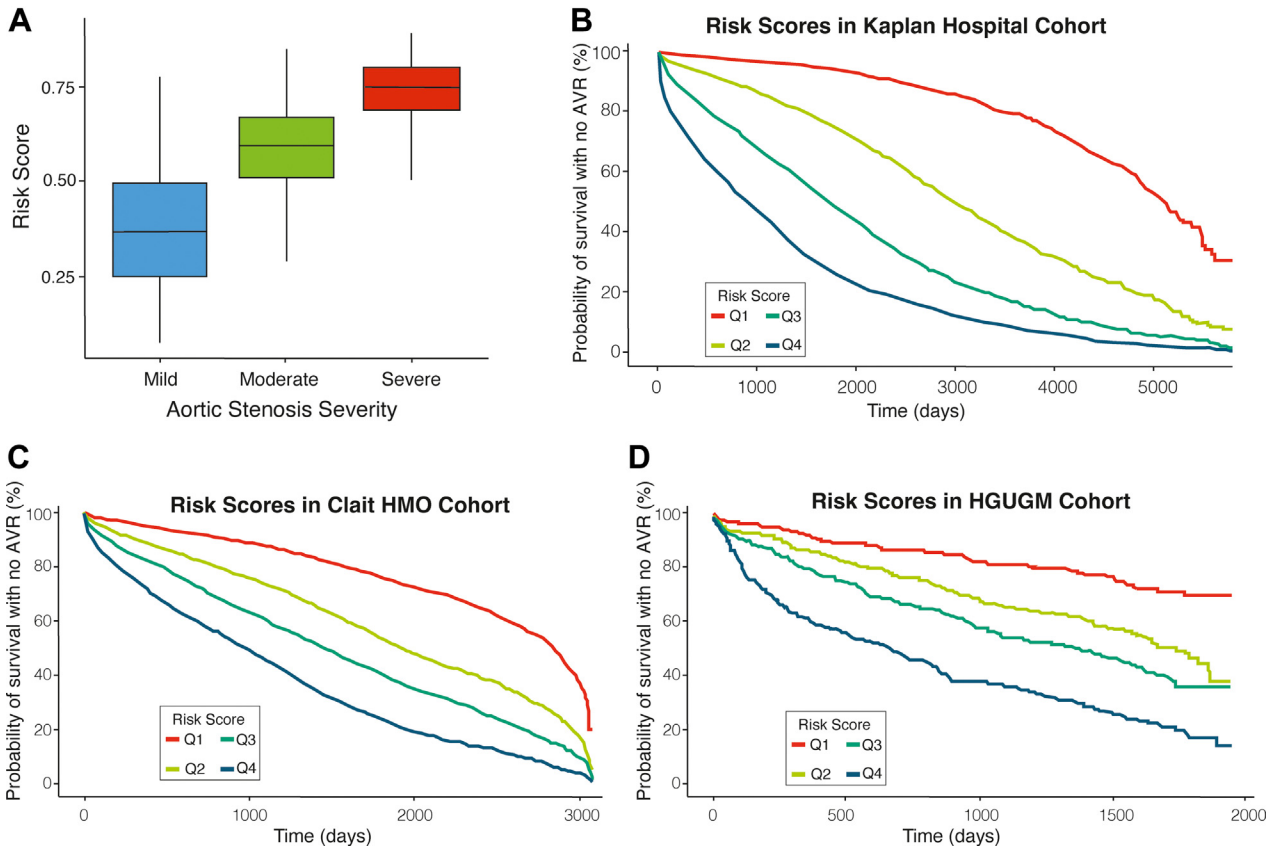
Several studies have employed ML techniques, primarily clustering and phenotyping, to enhance risk stratification for AS patients. Notably, the patient population in the primary registry significantly surpasses that of these previous studies. Bootstrap Lasso Regression model was used to develop ASteRisk score, which outperformed the classical clinical score, in patients with moderate and severe AS.[18] The accuracy of both the Cox proportional-hazards model and the RSF model in our study was higher, likely owing to the larger registry size, the utilization of advanced ML methods, and the incorporation of a greater number of variables. We included patients with mild AS as well, as done by Sengupta et al.[6] In their research, an unsupervised topological data analysis method was employed to identify and classify AS patient groups with similar phenotypes. Their ML model demonstrated superiority over conventional AS severity grading in terms of accuracy and prognostic capability, using only echocardiographic parameters. Notably, they found that 9% of patients initially classified as having nonsevere AS according to standard criteria were reclassified into the high-severity ML group. In our study, some patients with mild AS also showed high-risk scores, while some

FIGURE 2 Validation of Model Performance in External Cohorts



Comparison of predictive performance (AUC over time) of RSF model and Cox proportional-hazards survival model to predict combined outcomes in both the validation cohorts of Clalit HMO (A) and HGUGM (B). AUC = area under the curve; HGUGM = Hospital General Universitario Gregorio Marañón; RSF = random survival forest.

FIGURE 3 Risk Scores

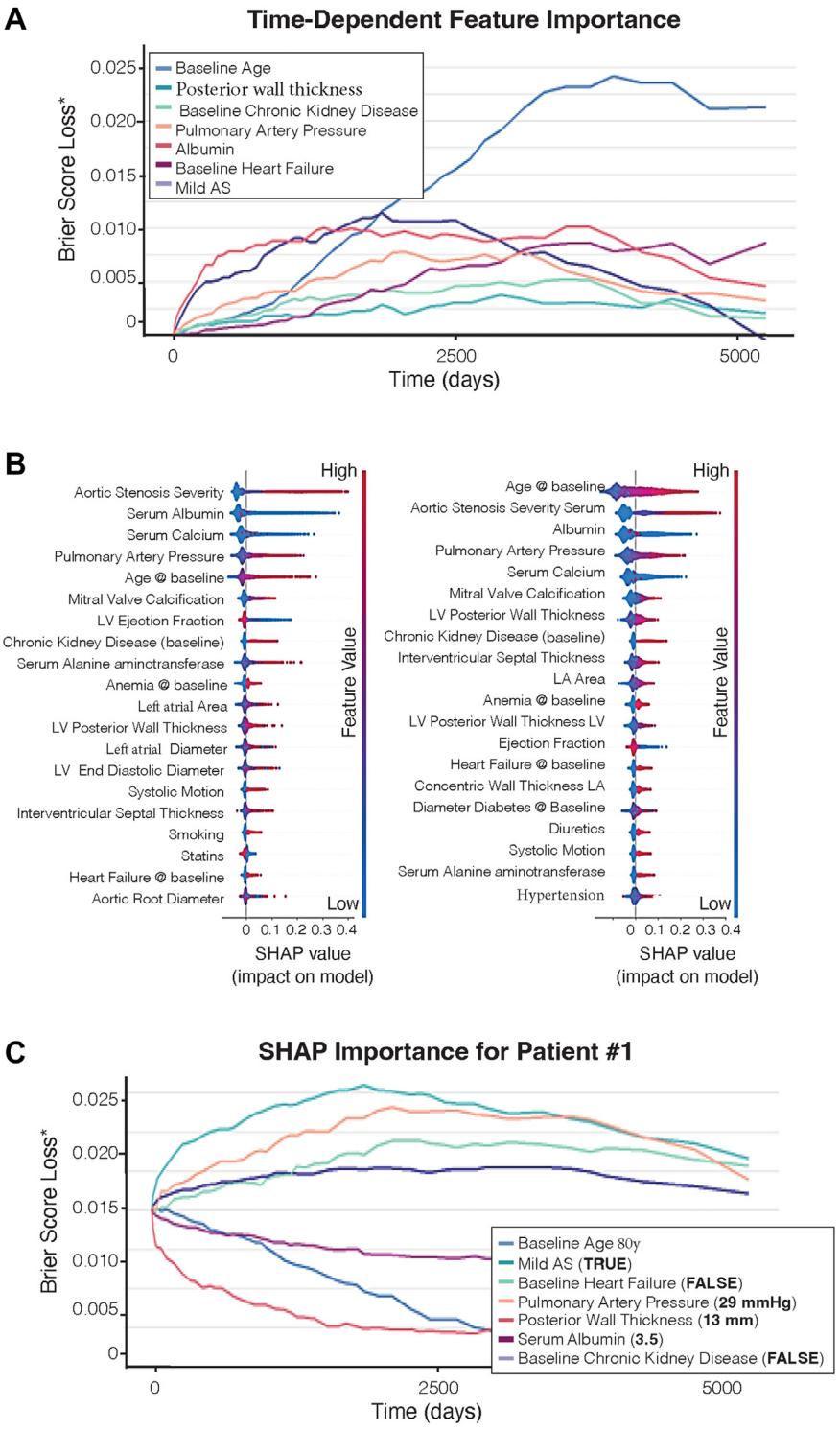


(A) Calculated relative risk score for patients with mild, moderate, and severe AS. Survival without AVR of patients according to model calculated risk scores quartiles, adjusted by age and sex in the original cohort (B), the Clalit cohort (C), and the HGUGM cohort (D). AVR = aortic valve replacement; HGUGM = Hospital General Universitario Gregorio Marañón; AS = aortic stenosis.

patients with severe AS, did not have the highest risk scores. Kwak et al[19] analyzed demographic, laboratory, and echocardiography parameters using unsupervised cluster analysis, revealing 3 distinct groups of patients with moderate and severe AS. The groups exhibited significant differences in terms of comorbidities, cardiac function, outcomes, and causes of death. Although Kwak et al[19] incorporated clinical data into the model, this data was limited.

In our study, important outcome predictors included AS severity, age, PAP, LV hypertrophy and function, left atrial size, albumin level, chronic kidney disease, anemia, and serum calcium. Samad et al[5] assessed clinical and echocardiographic data to predict patient survival, not restricted to AS patients, for up to 5 years. Random forest model proved to be effective for predicting 5-year survival, with age and tricuspid regurgitation velocity, representing PAP, emerging as the most important variables. In our cohort, age and PAP also stood out as significant variables, alongside AS severity, LV-related factors, and comorbidities, which align with findings from other studies. Serum albumin is a prognostic biomarker in various cardiovascular diseases, including heart failure, coronary artery disease, and stroke.[20] The reduced albumin level may be due to nutritional deficiency, inflammation, and renal or liver disease. The results identify serum albumin as a prognostic risk factor for patients with AS. Serum calcium, a significant variable in our model, has prompted contradictory reports regarding its role in AS.[21,22] The significance of serum calcium may be influenced by low albumin levels. While supplemental calcium was not a highly significant variable in our model, there have been reports suggesting a potential poor outcome association between supplemental calcium and vitamin D in patients with mild to moderate AS and low serum calcium levels.

FIGURE 4 Feature Importance



(A) Time-dependent Brier score for right-censored data, presenting the time-dependent importance of the various variables. Brier score estimates the average impact of each variable on the model's predictive performance. (B) Top variables for predicting both 1-year (left) and 5-year (right) outcomes. Importance scores are scaled to show relative importance according to variable value. Red color represents a high variable value, while blue color represents a low variable value. (C) An example of SHAP analysis results, showing patients' specific outcome predictors over time. AS = aortic stenosis; LV, left ventricle; LA = left atrial; SHAP = SHapley Additive exPlanations.

The developed algorithm exhibited good diagnostic accuracy. When tested on external cohorts, the diagnostic accuracy remained good, albeit slightly lower compared to the primary cohort. Such a difference in accuracy between the development and validation cohorts is a common occurrence in ML. It is important to bear in mind that a ML algorithm's knowledge is solely derived from the development dataset, and it is expected to exhibit similar or lower performance when applied to external datasets.[23] Additionally, variations in patient populations, particularly differences in the frequency of comorbidities and missing variables in both external datasets, can contribute to this variance in accuracy. Outcome-significant variables, such as PAP, were not available in the Clalit HMO database and were present only in about 70% of patients in the HGUGM validation cohort. Mitral annular calcification data were not available in both validation cohorts. However, it is worth emphasizing that the variables used in the model, including PAP, mitral annular description, and others, are "real-world" data typically found in the medical records of elderly patients.

Pressure overload and myocardial hypertrophy may play a significant role in a patient's prognosis, particularly in cases of severe and likely moderate AS, and the hemodynamic consequences are accompanied by ongoing cardiac damage over time. Recently, a staging classification has been proposed, which enhances risk stratification for patients with severe AS.[24] Staging includes LV remodeling, left atrial remodeling and mitral regurgitation, PAP, and right ventricular changes. Many of the highly important outcome variables in our model align with the significant determinants of this staging system. Our results extrapolate this staging to patients with lesser AS severity and add extracardiac variables. Left ventricular hypertrophy, whether entirely related to AS or not, serves as a negative predictive variable, as do other factors such as a dilated left atrium or elevated PAP. Additionally, the model explores other factors that are significant for patient outcomes. Variables such as serum albumin, renal function, and potentially liver function, along with anemia, carry significance in these patients. These factors should be considered as valuable variables for staging. A deeper understanding of the pathophysiological complexity and heterogeneity among AS patients may provide new insights that inform disease management, expand therapeutic strategies, and ultimately optimize patient outcomes.[25]

Using the ML model to assess patients' risk, potentially allows us to make informed decisions about whether a patient requires frequent follow-up and comorbidity management. Our model can contribute valuable information to enhance personalized approaches for valvular follow-up and echocardiographic examination timing. It enables to monitor and address specific conditions such as renal function, albumin levels (indicating nutritional status), and signs of heart failure. For certain patients, therapies targeting heart failure and renal failure, such as sodium-glucose co-transporter-2 inhibitors or spironolactone, may also be beneficial. Currently, there is no indication to perform valve intervention in patients with moderate AS or asymptomatic patients with severe AS. Several randomized clinical trials are being conducted and are very much awaited to answer these important clinical questions. While our study is observational and does not provide insights into the time for intervention, it suggests the possibility of AVR when a patient with moderate AS and high-risk score accompanies left ventricular hypertrophy, a dilated left atrium, and elevated PAP, since these variables are associated with a poor prognosis.

STUDY LIMITATIONS. We did not include longitudinal, repeated measurement information among the predictors. Due to the registry's inclusion criteria, which required only one echocardiogram, this parameter lacked validity for a significant proportion of patients. In addition, validation was conducted using retrospective data from the cohorts. Prospective, long-term validation of a large number of patients would require an extensive amount of time for inclusion and follow-up. Additionally, our dataset lacks information on patients' symptoms and biomarkers, such as B-type natriuretic peptide (BNP). The clinical significance of symptoms and NYHA functional class in patients with mild AS and even moderate AS is not known. In addition, symptoms in this high-risk and elderly population can be nonspecific. BNP levels were shown to have prognostic significance in patients with AS, mainly severe AS. Unfortunately, BNP testing was not conducted routinely in all AS patients during the study period. Left ventricular and left atrial strain are also important factors in AS progression. This data was not routinely collected during the study period.

CONCLUSIONS

Using advanced ML methods, we have developed an algorithm that accurately predicts patient outcomes across all grades of AS while identifying crucial prognosis-related variables. This model assesses patients' risk for adverse events, considering the time-dependent impact of each variable, and offers the

potential to guide risk factor modification for improved prognosis. Future enhancements to the model will involve the inclusion of additional variables, such as LV and left atrial strain, as well as the assessment of cumulative exposure to variable states and therapies during follow-up.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Research data was retrieved from the central data warehouse using the Clalit Health Services Secure Data Sharing Platform powered by MDClone (https://www.mdclone.com). This study was partially supported by the Instituto de Salud Carlos III (PI20/00587, to Dr Martinez-Legazpi) and by the EU—European Regional Development Fund. The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Sara Shimoni, Kaplan Medical Center, Pasternak Street 1, Rehovot, Israel. E-mail: Sara_s@clalit.org.il.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Aortic valve stenosis of any degree is associated with poor outcomes. However, AVR is indicated in patients with severe symptomatic AS or severe AS with reduced left ventricular systolic function.

COMPETENCY IN PATIENT CARE: Using the ML method, it is possible to assess patients' risk and identify variables that are associated with adverse outcomes in patients with all grades of AS. The algorithm offers the potential to guide risk factor modification for improved prognosis.

TRANSLATIONAL OUTLOOK: This study was based on a large cohort of patients with AS and was validated in 2 large independent cohorts. A prospective validation will improve the algorithm.

REFERENCES

1. d'Arcy JL, Coffey S, Loudon MA, et al. Large-scale community echocardiographic screening reveals a major burden of undiagnosed valvular heart disease in older people: the OxVALVE Population Cohort Study. *Eur Heart J*. 2016;37:3515–3522.

2. Investigators UTT, Toff WD, Hildick-Smith D, et al. Effect of transcatheter aortic valve Implantation vs surgical aortic valve replacement on all-cause mortality in patients with aortic stenosis: a randomized clinical trial. *JAMA*. 2022;327:1875–1887.

3. Strom JB, Playford D, Stewart S, et al. Increasing risk of mortality across the spectrum of aortic stenosis is independent of comorbidity & treatment: an international, parallel cohort study of 248,464 patients. *PLoS One*. 2022;17:e0268580.

4. Alcon B, Martinez-Legazpi P, Stewart S, et al. Transvalvular jet velocity, aortic valve area, mortality, and cardiovascular outcomes. *Eur Heart J Cardiovasc Imaging*. 2022;23:601–612.

5. Samad MD, Ulloa A, Wehner GJ, et al. Predicting survival from large echocardiography and electronic Health record datasets: optimization with machine learning. *JACC Cardiovasc Imaging*. 2019;12:681–689.

6. Sengupta PP, Shrestha S, Kagiyama N, et al. A machine-learning framework to identify distinct phenotypes of aortic stenosis severity. *JACC Cardiovasc Imaging*. 2021;14:1707–1720.

7. Baumgartner H, Hung J, Bermejo J, et al. Echocardiographic assessment of valve stenosis: EAE/ASE recommendations for clinical practice. *Eur J Echocardiogr*. 2009;10:1–25.

8. Baumgartner H, Hung J, Bermejo J, et al. Recommendations on the echocardiographic assessment of aortic valve stenosis: a focused update from the European Association of Cardiovascular Imaging and the American Society of Echocardiography. *Eur Heart J Cardiovasc Imaging*. 2017;18:254–275.

9. Pibarot P, Dumesnil JG. Aortic stenosis suspected to be severe despite low gradients. *Circ Cardiovasc Imaging*. 2014;7:545–551.

10. Lancellotti P, Moura L, Pierard LA, et al. European Association of Echocardiography recommendations for the assessment of valvular regurgitation. Part 2: mitral and tricuspid regurgitation (native valve disease). *Eur J Echocardiogr*. 2010;11:307–332.

11. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16:233–270.

12. Codolosa JN, Koshkelashvili N, Alnabelsi T, Goykhman I, Romero-Corral A, Pressman GS. Effect of mitral annular calcium on left ventricular diastolic parameters. *Am J Cardiol*. 2016;117:847–852.

13. [preprint] Anthony Devaux (BPH) CHB, Robin Genuer (SISTM, BPH), Cécile Proust-Lima (BPH). Random survival forests with multivariate longitudinal endogenous covariates. *arXiv*. 2023. https://doi.org/10.48550/arXiv.2208.05801

14. Chen Y, Jia Z, Mercola D, Xie X. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. *Comput Math Methods Med*. 2013;2013:873595.

15. Stewart S, Chan YK, Playford D, Strange GA, Investigators N. Incident aortic stenosis in 49 449 men and 42 229 women investigated with routine echocardiography. *Heart*. 2022;108:875–881.

16. Messika-Zeitoun D, Baumgartner H, Burwash IG, et al. Unmet needs in valvular heart disease. *Eur Heart J*. 2023;44:1862–1873.

17. Madakkatel I, Zhou A, McDonnell MD, Hypponen E. Combining machine learning and conventional statistical approaches for risk factor discovery in a large cohort study. *Sci Rep*. 2021;11:22997.

18. Namasivayam M, Myers PD, Guttag JV, et al. Predicting outcomes in patients with aortic stenosis using machine learning: the Aortic Stenosis Risk (ASteRisk) score. *Open Heart*. 2022;9:e001990.

19. Kwak S, Lee Y, Ko T, et al. Unsupervised cluster analysis of patients with aortic stenosis reveals distinct population with different phenotypes and outcomes. *Circ Cardiovasc Imaging*. 2020;13:e009707.

20. Chien SC, Chen CY, Lin CF, Yeh HI. Critical appraisal of the role of serum albumin in cardiovascular disease. *Biomark Res*. 2017;5:31.

21. Xia C, Lei W, Hu Y, Yang H, Zeng X, Chen M. Association of serum levels of calcium, phosphate, and vitamin D with risk of developing aortic stenosis: the UK Biobank cohort. *Eur J Preventive Cardiol*. 2022;29:1520–1528.

22. Wald DS, Bangash FA, Morris JK, Wald NJ. Mortality from aortic stenosis: prospective study of serum calcium and phosphate. *J Internal Med*. 2017;281:407-411.

23. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. *Radiol Artificial Intelligence*. 2022;4:e210064.

24. Genereux P, Pibarot P, Redfors B, et al. Staging classification of aortic stenosis based on the extent of cardiac damage. *Eur Heart J*. 2017;38:3351-3358.

25. Gayle KA, Lindman BR. Uncovering the phenotypic heterogeneity of patients with aortic stenosis: a path to new insights? *Circul Cardiovasc Imaging*. 2020;13:e010786.

KEY WORDS aortic valve stenosis, aortic stenosis prognosis, machine learning, random survival forest model

APPENDIX For supplemental methods, tables, and figures, please see the online version of this paper.