JACC: CARDIOVASCULAR IMAGING
© 2023 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
PUBLISHED BY ELSEVIER

VOL. 16, NO. 6, 2023

ORIGINAL RESEARCH

Machine Learning to Optimize the Echocardiographic Follow-Up of Aortic Stenosis

Antonio Sánchez-Puente, BSc, PhD,[a,b,*] P. Ignacio Dorado-Díaz, BE, PhD,[a,b,*] Jesús Sampedro-Gómez, BE,[a,b] Javier Bermejo, MD, PhD,[b,c] Pablo Martinez-Legazpi, BE, PhD,[d] Francisco Fernández-Avilés, MD, PhD,[b,c] Javier Sánchez-González, BE, PhD,[e] Candelas Pérez del Villar, MD, PhD,[a,b] Víctor Vicente-Palacios, BE, PhD,[e,†] Pedro L. Sanchez, MD, PhD[a,b,†]

ABSTRACT

BACKGROUND Disease progression in patients with mild-to-moderate aortic stenosis is heterogenous and requires periodic echocardiographic examinations to evaluate severity.

OBJECTIVES This study sought to explore the use of machine learning to optimize aortic stenosis echocardiographic surveillance automatically.

METHODS The study investigators trained, validated, and externally applied a machine learning model to predict whether a patient with mild-to-moderate aortic stenosis will develop severe valvular disease at 1, 2, or 3 years. Demographic and echocardiographic patient data to develop the model were obtained from a tertiary hospital consisting of 4,633 echocardiograms from 1,638 consecutive patients. The external cohort was obtained from an independent tertiary hospital, consisting of 4,531 echocardiograms from 1,533 patients. Echocardiographic surveillance timing results were compared with the European and American guidelines echocardiographic follow-up recommendations.

RESULTS In internal validation, the model discriminated severe from nonsevere aortic stenosis development with an area under the receiver-operating characteristic curve (AUC-ROC) of 0.90, 0.92, and 0.92 for the 1-, 2-, or 3-year interval, respectively. In external application, the model showed an AUC-ROC of 0.85, 0.85, and 0.85, for the 1-, 2-, or 3-year interval. A simulated application of the model in the external validation cohort resulted in savings of 49% and 13% of unnecessary echocardiographic examinations per year compared with European and American guideline recommendations, respectively.

CONCLUSIONS Machine learning provides real-time, automated, personalized timing of next echocardiographic follow-up examination for patients with mild-to-moderate aortic stenosis. Compared with European and American guidelines, the model reduces the number of patient examinations. (J Am Coll Cardiol Img 2023;16:733–744)
© 2023 by the American College of Cardiology Foundation.

From the [a]Cardiology Service, Salamanca University Hospital, Biomedical Research Institute of Salamanca (IBSAL), Department of Medicine, University of Salamanca, Salamanca, Spain; [b]Spanish Cardiovascular Network (CIBERCV), Carlos III Health Institute, Spain; [c]Cardiology Service, Gregorio Marañón University Hospital, Gregorio Marañón Health Research Institute (IISGM), Faculty of Medicine, Complutense University, Madrid, Spain; [d]Department of Mathematical Physics and Fluids, Faculty of Sciences, National University of Distance Education (UNED) and CIBERCV, Madrid, Spain; and [e]Philips Healthcare, Madrid, Spain. *Drs Sánchez-Puente and Dorado-Díaz contributed equally to this paper as co-first authors. †Drs Vicente-Palacios and Sanchez contributed equally to this paper as co-senior authors.

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 April 5, 2022; revised manuscript received November 17, 2022, accepted December 2, 2022.

ISSN 1936-878X/$36.00 https://doi.org/10.1016/j.jcmg.2022.12.008

ABBREVIATIONS AND ACRONYMS

AS = aortic stenosis

AUC-PR = area under the precision-recall curve

AUC-ROC = area under the receiver-operating characteristic curve

ML = machine learning

Aortic stenosis (AS) is the most prevalent valve disease and an important cause of morbidity and mortality.[1] A patient with mild or moderate (stage B) AS may eventually have severe valve obstruction (stage C) and therefore will be at increased risk of heart failure, syncope, angina, or death.[2] Once symptoms appear, aortic valve replacement is the only treatment capable of changing the natural history of the disease and improving prognosis. The rate of hemodynamic progression is heterogenous, and the onset of clinical symptoms is variable. For these reasons, it is recommended that all patients with known AS should have periodic clinical and echocardiographic evaluations at regular intervals, and the interval length depends on the patient's valve stenosis severity.[3-7] However, there is limited knowledge on what constitutes an optimal echocardiographic follow-up interval. The current surveillance schemes are different among geographic regions, and all have Level of Evidence: C. American guidelines suggest echocardiography examination every 6 to 12 months for asymptomatic severe AS, every 1 to 2 years for moderate AS, and every 3 to 5 years for mild AS.[3,6] These intervals are different in European guidelines, which recommend follow-up echocardiography every 6 months in cases of severe AS and once yearly for mild and moderate AS.[4,7]

There is growing interest in precision medicine techniques that can deliver personalized patient care. In this sense, machine learning (ML) techniques are able to predict patient outcomes reliably by using individual patient observations.[8,9] Although ML models applied to AS can be found in the published reports,[10-12] none of these reports focus on follow-up, nor have we found examples that apply artificial intelligence methodology when analyzing the optimal time interval to perform an imaging follow-up examination for a chronic condition. However, there are good examples of algorithms capable of real-time monitoring follow-up recommendations in radiology reports.[13,14] Notwithstanding, the ML classifiers used in our proposal are similar to other cardiovascular disease risk prediction models showing higher performance compared with conventional statistical methods.[15]

The aim of this study was to develop 3 different supervised ML models fed by echocardiographic measurements of an AS patient and predict whether severe AS would develop in this patient, as assessed by follow-up echocardiography at 1, 2, or 3 years after baseline examination. These models were applied in a system that recommends the time for the next follow-up in new AS patients.

METHODS

The steps followed can be divided into data preparation, model training, model evaluation, and model application, as summarized in **Figure 1**.

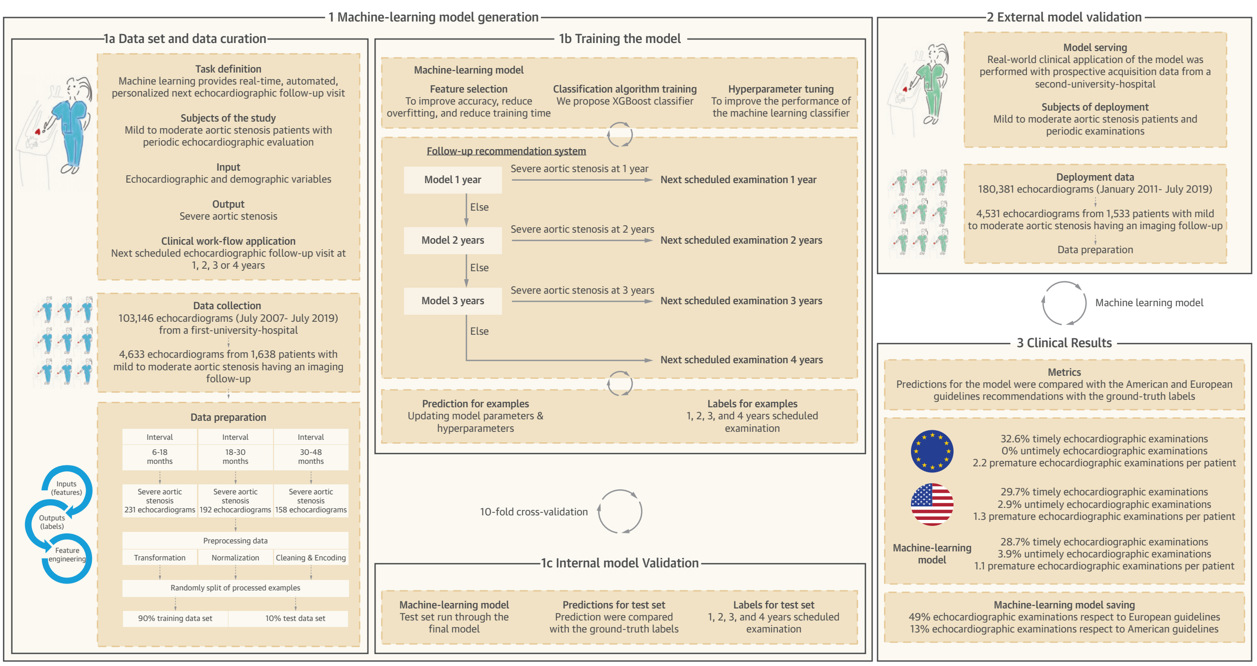
This study followed the PRIME (Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation) checklist[16] for the development of ML models in cardiology (Supplemental Table 1).

MACHINE LEARNING MODEL GENERATION. Training cohort and data curation. The demographic and echocardiographic data used to develop the model were extracted from the archive system (IntelliSpace Cardiovascular, Philips Healthcare) of the tertiary University Hospital of Salamanca, Spain, and comorbidities were extracted from the hospital electronic health records. The echocardiographic database gathers raw tabular data from a total of 103,146 echocardiographic studies from 60,657 patients screened between July 2007 and July 2019, from all the multibrand echocardiographers of the cardiology department (Philips, General Electric, and Toshiba). To develop the ML model, we included all patients with initial mild-to-moderate AS who had at least 2 periodic imaging assessments. Doppler echocardiographic transaortic peak velocity was used for defining severity of AS that was graded as mild from 2.0 to 2.9 m/s, moderate from 3 to 3.9 m/s, or severe when peak velocity was ≥4 m/s.[3,6] The intervals between visits were categorized as 1 year if the next follow-up echocardiography was performed in the 6- to 18-month interval, a 2 years in the 18- to 30-month interval, and as 3 years in the 30- to 42-month interval. Target labels for algorithm training were defined by whether the patient had severe AS or not on each of those intervals.

This training data set was composed of 92 demographic and echocardiographic variables. Only variables with at least 30% of nonmissing values were considered. All echocardiographic variables were continuous. Missing values in continuous variables were filled with the mean value obtained for the rest of the patients. Comorbidities were considered negative if not coded.

Training the machine learning model. The training process involved feature selection, classification algorithm training, and hyperparameter tuning. Before training the actual model, a feature selection operation was performed by combining multivariate analysis of variance and biplot methods,[17] thus allowing

FIGURE 1 Machine Learning Model Workflow



Overview of the phases followed to build and apply the machine learning model and matching learning prediction results applied to clinical decision making.

simultaneous hyperspatial representations of subjects (as points) and variables analyzed (as vectors), along with their binary characterization (severe vs nonsevere AS). By projecting the subjects of each category on the variables, we were able to assign a discriminant value to each of them. These values were used to establish a ranking (Supplemental Figure 1). From the feature selection analysis, a final set of 10 variables was selected to develop the model, listed according to the variables' importance: peak aortic jet velocity, mean aortic velocity, aortic velocity time integral, patient age, left ventricular mass, slope of deceleration of the mitral E wave, left ventricular ejection fraction, left ventricular stroke volume, mean left ventricular outflow tract velocity, and left ventricular end-diastolic volume.

Model development codes were written in Python and made use of the open code library scikit-learn.[18] To predict AS echocardiographic follow-up, we proposed an XGBoost (open source software) classifier because of its high efficiency and versatility.[9] A classification model was chosen in favor of a regression model because the prediction error was larger for regression models when trying to give more specific temporal data as output. The choice of the XGBoost algorithm was made because of its proven better performance,[19] and it was based on the technical merits of improving and updating these models in the future.

We tuned the hyperparameters to improve the performance of the ML classifiers (Supplemental Table 2). Automated hyperparameter selection was done according to best performance in stratified 10-fold cross-validation measured as the area under the receiver-operating characteristic curve (AUC-ROC), later also used to evaluate our final models.

Internal model validation. The individual ML classifiers were internally evaluated with a 10-fold cross-validation scheme with 10 repetitions.[20] Because training of the models includes a hyperparameter tuning step that also performs its own cross-validation step, this resulted in nested cross-validations. Predictions for severe AS development at 1, 2, and 3 years for the test set were compared with the ground-truth labels.

Metrics and follow-up recommendations. The ROC and the precision-recall (PR) curve analyses were used to assess the predictive capacity of each individual ML model.

To provide an optimal echocardiographic follow-up window, we used the probability function of the classifier together with a cutoff threshold. Therefore,

JACC: CARDIOVASCULAR IMAGING, VOL. 16, NO. 6, 2023
JUNE 2023:733-744

FIGURE 2 Patients' Flow Diagram for Both Data Sets

First-university-hospital population
(model development)

103,146 echocardiograms from 60,657 patients eligible between July 2007 and July 2019

98,513 echocardiograms and 59,019 patients excluded
- 88,191 echocardiograms and 53,024 patients with non-aortic stenosis
- 3,840 echocardiograms and 1,051 patients with prosthetic aortic valve (or stress echocardiogram performed)
- 973 echocardiograms following a first severe aortic stenosis assessment
- 55 echocardiograms and 18 patients with aortic stenosis inconsistency measurement
- 217 echocardiograms followed by another study within one month
- 4,735 echocardiograms and 4,735 patients with aortic stenosis but unique evaluation
- 502 echocardiograms and 191 patients with no data regarding aortic stenosis severity

4,633 echocardiograms from 1,638 patients with initial mild to moderate aortic stenosis
- 1,638 echocardiograms corresponding to initial evaluation
- 2,995 echocardiograms corresponding to periodic monitoring

Second-university-hospital population
(model deployment)

180,381 echocardiograms from 89,802 patients eligible between January 2011 and July 2019

175,850 echocardiograms and 88,269 patients excluded
- 159,623 echocardiograms and 81,221 patients with non-aortic stenosis
- 10,815 echocardiograms and 3,242 patients with prosthetic aortic valve (or stress echocardiogram performed)
- 1,138 echocardiograms following a first severe aortic stenosis assessment
- 81 echocardiograms and 19 patients with aortic stenosis inconsistency measurement
- 195 echocardiograms followed by another study within one month
- 3,678 echocardiograms and 3,678 patients with aortic stenosis but unique evaluation
- 320 echocardiograms and 109 patients with no data regarding aortic stenosis severity

4,531 echocardiograms from 1,533 patients with initial mild to moderate aortic stenosis
- 1,533 echocardiograms corresponding to initial evaluation
- 2,998 echocardiograms corresponding to periodic monitoring

The training cohort where the machine learning model was developed and the external cohort where the machine learning model was deployed.

a patient who surpassed the risk threshold for severe AS development at 1 year was assigned to a 1-year follow-up. We followed same strategy to assign 2- and 3-year follow-ups. If the risk was below all model thresholds, the patient was assigned to a 4-year follow-up screening. The cutoff thresholds were chosen after performing a grid search for the combination that results in fewer echocardiographic examinations per year, on the condition of having at most the same number of late echocardiographic examinations than the American guidelines, when evaluated on the results of internal validation.

The classification performance at particular cutoff thresholds was also evaluated according to its sensitivity, specificity, precision, and negative predictive value, in the group not scheduled for a follow-up by the previous model and threshold.

EXTERNAL APPLICATION. The model and cutoff values were evaluated in an external cohort from a second and different tertiary university hospital in Madrid, Spain. Data were obtained from a series of 180,381 unselected consecutive echocardiograms from 89,802 patients screened between January 2011 to July 2019. Outcome data of the composite endpoint of aortic valve replacement (either surgical or percutaneous) or any-cause mortality were recorded from the electronic medical records for this cohort.

Echocardiographic surveillance recommendations on the basis of the whole ML system (applying the 3 models as described earlier), the European guidelines,[4] and the American guidelines[6] were compared with the ground-truth labels. Echocardiographic follow-ups scheduled by each system were labeled as follows: *premature,* if the patient did not develop severe AS within 6 months of the predicted interval; *timely,* if the scheduled follow-up was within 6 months of the diagnosis of severe AS; and *untimely,* if the patient developed severe AS more than 6 months before the recommended follow-up. For the efficiency calculation, we restricted the analysis to the 856 patients (of 1,533) whose AS severity status was assessed yearly during the 3 years following after the baseline visit. A total of 100 bootstrap resamples were used to provide CIs for these measurements and McNemar's test was used for significance testing.

The performance of the individual models in subgroups of clinical interest was analyzed in this external application in terms of the AUC-ROC and area under the PR curve (AUC-PR) curves.

Institutional approvals to undertake the study were provided by the Local Ethics Committees of both centers (approvals 2018/10/127 in Salamanca and 228/17 in Madrid), which exempted the need for patient informed consent because the study required no modification in standard clinical practice. All data sets were anonymously analyzed, and the study was performed following current recommendations of the Declaration of Helsinki.

RESULTS

MACHINE LEARNING MODEL DEVELOPMENT. From the training cohort data set (flowchart in **Figure 2**), we identified 1,638 patients with initial mild or moderate AS who underwent 4,633 echocardiographic studies; 2,995 of these studies corresponded to periodic monitoring different from the initial evaluation.

Patients had a mean age of 73 ± 11 years at the initial evaluation, 52.4% were men, and 1,124 (68.6%) presented with mild AS and 514 (31.4%) with moderate AS. Comorbidities recorded were hypertension (67.7%), dyslipidemia (54.8%), diabetes (29.0%), a history of tobacco use (22.0%), current tobacco use (4.1%), and chronic kidney disease with glomerular filtration rate lower than 60 mL/min/1.73 m^2 (3.7%). Each patient had an average of 1.8 ± 1.3 periodic examinations after the first echocardiographic study, and the average time between examinations was 2.0 ± 1.5 years. Duration of follow-up was 3.6 ± 2.4 years, ranging from 0.1 to 11.3 years. Normalized for length of follow-up and expressed as an annual rate of change, aortic jet velocity increased by 0.18 ± 0.64 m/s per year, and mean gradient increased by 3.15 ± 14.1 mm Hg per year. Diagnosis of severe AS was made during follow-up in 581 (19.4%) of the 2,995 periodic echocardiographic examinations: in 231 of these studies (39.8%) this change was observed over the first year of follow-up; it was seen in 192 (33.0%) over the second year and in 158 (27.2%) over the third year.

TABLE 1 AUC-ROC for the Developed ML Models in the Internal Validation

ML Model	Algorithm	AUC-ROC
1 y	L2-regularized logistic regression	0.90 (0.89-0.92)
	Random Forest	0.90 (0.88-0.92)
	XGBoost	0.90 (0.88-0.92)
2 y	L2-regularized logistic regression	0.92 (0.91-0.93)
	Random Forest	0.92 (0.91-0.93)
	XGBoost	0.92 (0.90-0.93)
3 y	L2-regularized logistic regression	0.92 (0.91-0.94)
	Random Forest	0.92 (0.91-0.93)
	XGBoost	0.92 (0.91-0.93)

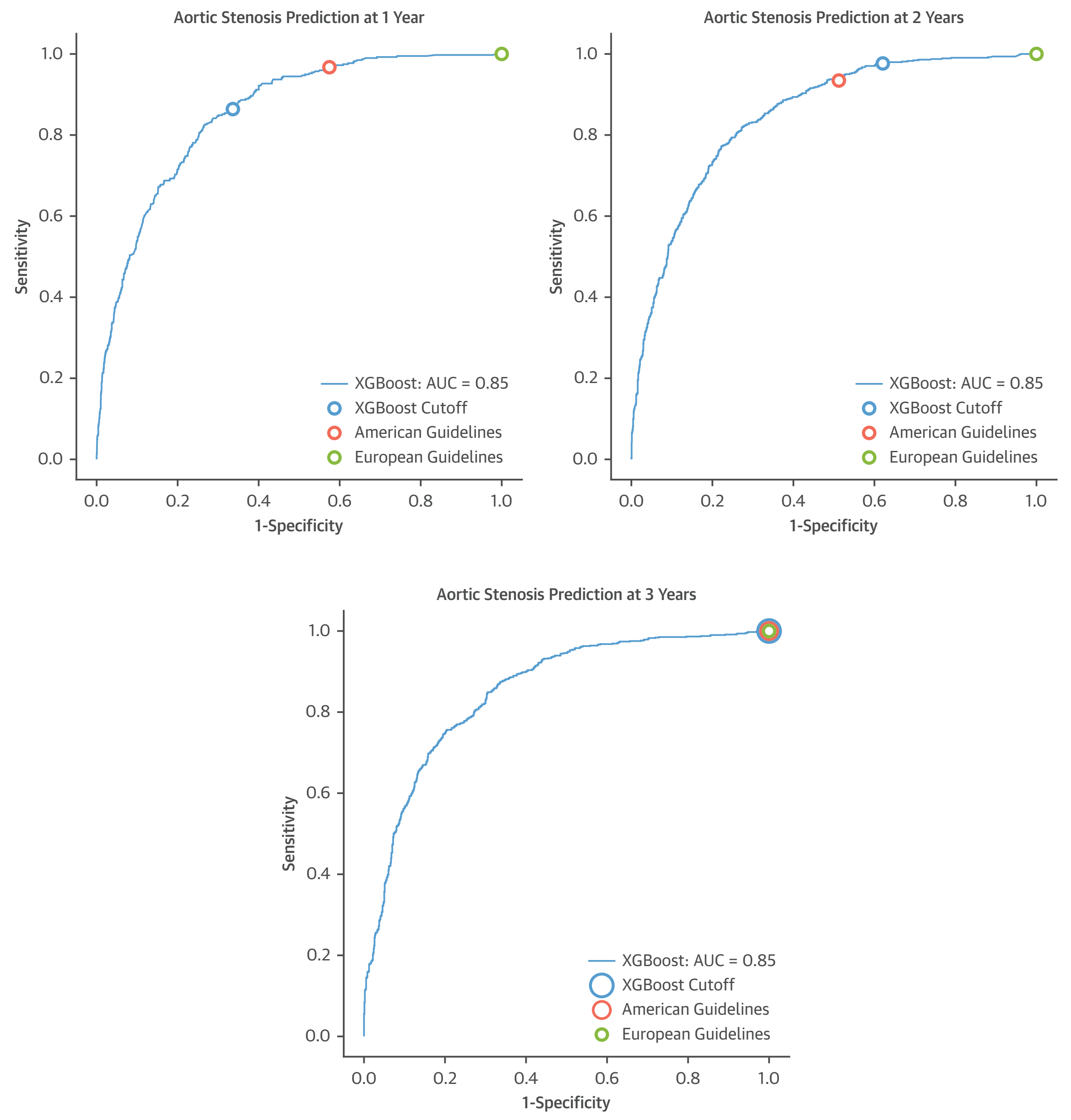
AUC-ROC = areas under the receiver-operating characteristic curves; ML = machine learning.

The prediction accuracy of the different ML algorithms under consideration is shown in **Table 1**, with a similar area under the ROC curve of 0.90 (95% CI: 0.88-0.92), 0.92 (95% CI: 0.90-0.93), and 0.92(95% CI: 0.91-0.93) for predicting severe AS at 1, 2, and 3 years, respectively. Demographic and echocardiographic variable information split by output label is shown in Supplemental Table 3. Of note, broadening the input into the model with comorbidities did not improve the accuracy of the ML algorithms (Supplemental Figure 2).

We chose operating thresholds for each ML model according to the internal validation results. The first model predicted severe AS at 1 year for all patients having an estimated risk >6.8% with a sensitivity of 90.8%, a specificity of 74.6%, precision of 24.2%, and a negative predictive value of 98.9%. In the same way, the second model predicted severe AS at 2 years for patients with an estimated risk >2.8% with a sensitivity of 70.9%, a specificity of 82.3%, precision of 11.4%, and a negative predictive value of 98.9%, in those patients not scheduled to have a follow-up at 1 year by the previous model. Finally, the third model predicted severe AS at 3 years in patients with an estimated risk >2.0% with a sensitivity of 77.1%, a specificity of 73.6%, precision of 9.3%, and a negative predictive value of 98.9%, in the patients not scheduled for follow-up by the previous 2 models.

JACC: CARDIOVASCULAR IMAGING, VOL. 16, NO. 6, 2023
JUNE 2023:733-744

FIGURE 3 Receiver-Operating-Characteristic Curves for Machine Learning Models



The area under the curve receiver-operating-characteristic (AUC) curves for each individual model of the machine learning follow-up recommendation system. The cutoff thresholds for each model are represented as points over the curve together with the corresponding point for European and American guidelines. The European guidelines have 0% specificity (they schedule a follow-up for all patients) but have 100% sensitivity (they always detect severe aortic stenosis in a timely manner). The machine learning model and American guidelines have only 0% specificity and 100% sensitivity for the 3-year model because they schedule a follow-up for 3 years whether the baseline stenosis is mild or moderate.

FIGURE 4 **Envisioned Clinical Use of the Machine Learning Prediction**

This figure used as an example a group of 100 patients from the external cohort data set for which ground truths (severe or nonsevere aortic stenosis) are known for each of the 3 surveillance years following the baseline examination. The severity of aortic stenosis is represented in different colors: **yellow** for mild, **orange** for moderate, and **red** for severe aortic stenosis. Each **circle** represents an echocardiographic examination: **black circles** are baseline examinations, **blue circles** are premature follow-ups, **green circles** are timely follow-ups, red **circles** are untimely follow-ups, and **dotted red circles** represent patients without scheduled follow-up in the third year whose next follow-up will be untimely. echos = echocardiographic follow-ups.

MACHINE LEARNING MODEL APPLICATION. The external application of the ML model and its comparison with current clinical practice guidelines were performed using the external cohort data (**Figure 2**). We identified 1,533 patients with initial mild-to-moderate AS who underwent 4,531 echocardiographic studies; 2,998 of these studies corresponded to periodic monitoring different from the initial evaluation.

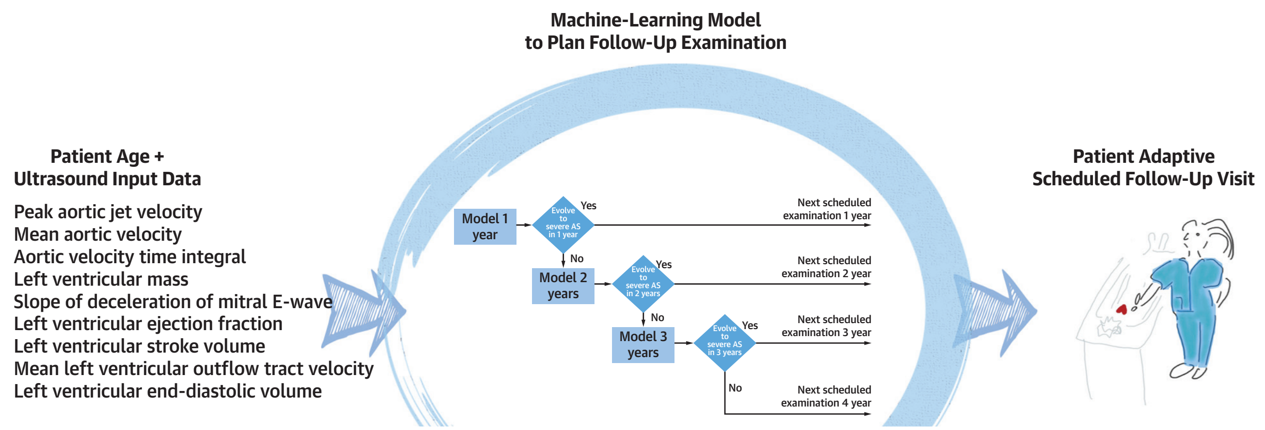
In this external cohort data set, patients had a mean age of 77 ± 11 years at initial evaluation, and 792 (51.7%) presented with mild AS and 741 (48.3%) with moderate AS. Each patient had an average of 2.0 ± 1.3 periodic examinations after their first echocardiographic study, the average time between examinations was 1.4 ± 1.0 years, and the duration of follow-up was 2.4 ± 1.7 years, ranging from 0.1 to 7.0 years. In this cohort, aortic jet peak velocity increased by 0.27 ± 0.68 m/s per year, and mean gradient increased by 4.3 ± 10.8 mm Hg per year. Diagnosis of severe AS was made during follow-up in 808 (27.0%) of the 2,998 periodic echocardiographic examinations: in 397 (49.1%), this change was observed over the first year of follow-up; in 244 (30.2%), it was seen over the second year; and in 167 (20.7%), it was noted over the third year of follow-up.

Performance of the model during external application obtained an AUC-ROC curve of 0.85 (95% CI: 0.84-0.87), 0.85 (95% CI: 0.84-0.87), and 0.85 (95% CI: 0.84-0.87) for predicting severe AS at 1, 2, and 3 years, respectively (**Figure 3**). Demographic and echocardiographic variable information split by output label is shown in Supplemental Table 4, and the contribution or the importance of each feature on the prediction of the model is shown in Supplemental Figure 3.

Specificity and sensitivity metrics obtained from the European and American guidelines lie over the model ROC curves. In this external validation cohort, a random estimator would have an AUC-PR curve of 0.14 for the 1-year model, 0.29 for the 2-year model, and 0.46 for the 3-year model, whereas our ML model showed an AUC-PR curve of 0.50, 0.69, and 0.82 at 1, 2, and 3 years, respectively, which is a substantial increase.

The efficiency of the ML follow-up recommendation and its comparison with European and American guidelines are shown in **Figure 4**. European guidelines

CENTRAL ILLUSTRATION Machine Learning for the Echocardiographic Follow-Up of Aortic Stenosis



Sánchez-Puente A, et al. J Am Coll Cardiol Img. 2023;16(6):733-744.

Machine learning methodology provides patient-adaptive follow-up echocardiographic examinations on the basis of patient age and echocardiographic measurements in patients with mild-to-moderate aortic stenosis (AS).

had 32.6% (95% CI: 29.6%-31.1%) of patients with *timely* follow-ups, 0% with *untimely* follow-ups, and 2.2 (95% CI: 2.1-2.3) *premature* follow-up echocardiographic examinations per patient (87% of the total echocardiograms). American guidelines had 29.7% (95% CI: 26.6%-32.3%) with *timely* follow-ups, 2.9% (95% CI: 2.0%-4.0%) with *untimely* follow-ups, and 1.3 (95% CI: 1.2-1.3) *premature* follow-up echocardiographic examinations per patient (80% of the total echocardiograms). The ML system had 28.7% (95% CI: 25.7%-31.1%) patients with *timely* follow-ups, 3.9% (95% CI: 2.9%-5.1%) with *untimely* follow-ups, and 1.1 (95% CI: 1.1-1.2) *premature* follow-up echocardiographic examinations per patient (77% of the total). The difference in untimely follow-ups between the ML system and the American guidelines was 1.0% (95% CI: −0.2% to 2.3%; $P = 0.18$), not statistically significant, and with respect to the European guidelines it was 3.9% (95% CI: 2.9%-5.1%; $P < 0.001$), statistically significant. When expressed as the percentage of unnecessary echocardiographic examinations per year, the ML system saved 13% and 49% of examinations when compared with American and European guidelines recommendations, respectively. Global performance for the ML system to predict the combined outcome endpoint of all-cause mortality or aortic valve replacement was moderate (AUC: 0.65 for the outcome prediction at 3 years; 95% CI: 0.63-0.67).

PERFORMANCE OF THE MODEL IN SELECTED SUBGROUPS. We evaluated the performance of the model for patients <55 years of age (who are likely to have aortic bicuspid valve AS), patients with low-flow low-gradient AS, patients with other concomitant valve disease, and patients with mild or moderate AS at evaluation in the external cohort. Discrimination of the model in patients with likely aortic bicuspid valve AS demonstrated superior performance as assessed by the AUC-ROC curve of 0.98, 1.00, and 0.89 and by the AUC-PR curve of 0.81, 0.95, and 0.83 for predicting severe AS at 1, 2, and 3 years, respectively (Supplemental Figure 4). Discrimination of the model in patients with low-flow low-gradient AS demonstrated good performance as assessed by the AUC-ROC curve of 0.87, 0.84, and 0.84 and by the AUC-PR curve of 0.82, 0.85, and 0.88 for predicting severe AS at 1, 2, and 3 years, respectively (Supplemental Figure 5). In contrast, discrimination of the model in patients with concomitant aortic regurgitation (Supplemental Figure 6) or concomitant mitral regurgitation (Supplemental Figure 7) did not improve the overall model performance in the external cohort. Interestingly, discrimination of the model in patients with mild or moderate AS demonstrated comparable performance (Supplemental Figure 8).

DISCUSSION

This study shows that patients with mild-to-moderate AS can be characterized using ML methods to schedule follow-up echocardiographic examinations that are tailored to the AS progression of specific

TABLE 2 Classification Performance of the ML Model Calculated for 3 Different Thresholds

	Timely Examinations, %	Untimely Examinations, %	Premature Examinations per Patient, n	Saving Examinations Compared With European Guidelines, %	Saving Examinations Compared With American Guidelines, %
ML model/thresholds 1	27.7	4.9	1.05	52	17
ML model/thresholds 2[a]	28.7	3.9	1.10	49	13
ML model/thresholds 3	31.3	1.3	1.46	34	−15
European guidelines	32.6	0	2.20	–	−41
American guidelines	29.7	2.9	1.27	41	–

[a]ML model/thresholds 2 used in Figure 4. Two additional sets of cutoff thresholds (model/thresholds 1 and 3) were chosen to be riskier and more conservative, respectively, to evaluate the possibilities of adapting the recommendations to each health care system.

Abbreviation as in Table 1.

individual patients (Central Illustration). To our knowledge, this investigation represents the first evidence regarding AS imaging surveillance and could set a basis for future guidelines with a Level of Evidence: C.

This study has implications for caregivers and patients. Using existing rigid follow-up intervals, AS patients receive unnecessary follow-up imaging examinations, which introduce unnecessary hospital workload and costs to the medical care system.[21,22] By contrast, the proposed ML model increases efficiency while also detecting patients who are likely to develop severe AS more rapidly.

To put numbers to these cost savings, if we applied our ML model to the European (741.4 million × 0.4%) and U.S. (327.2 million × 0.4%) AS groups,[1] we estimate we could save 180,000 to 150,000 echocardiographic examinations per year, and the consequent cost saving (in U.S. dollars) would be $83,160,000 and $69,300,000, respectively (calculation estimated with a fee for a formal hospital-based echocardiogram of $462).[23] If we considered the data from the external cohort, the cost to perform a timely examination was reduced with our ML model by between $353,589 per year with respect to the most conservative European guideline and $95,172 per year with respect to the American guideline.

Importantly, the ML model allows us to individualize its thresholds (Table 2) depending on caregivers (eg, U.S. access to health care services is lagging behind Europe in general, or cost of health care is higher in the United States compared with Europe). For example, policymakers would have the option to use an ML model reducing use of echocardiographic surveillance to 1.05 premature follow-ups per patient (saving 54% and 17% examinations when compared with European and American guidelines, respectively) at increasing the number of untimely examinations up to 5%; or they would have the option of using an ML model reducing the number of untimely examinations down to 1% at increasing 1.5 premature follow-ups (still saving 34% examinations when compared with European guidelines but increasing 15% examinations when compared with American guidelines).

Little is known about appropriate echocardiography intervals for repeat imaging testing in specific cardiac clinical settings,[24,25] even in valvular heart disease.[3] The absence of relevant research for AS patients has forced the guidelines to rely on the consensus opinion of experts,[4,6] with different consideration among guidelines. Our ML system adds value to the current Level of Evidence: C guideline recommendations because it is able to replicate the results, and, in addition, it enables the possibility of tailoring the classification thresholds to schedule follow-up recommendations depending on the resource availability of each health care system. Artificial intelligence models can provide information on Level of Evidence: C recommendations, which still represent 41.5% among recommendations in major cardiovascular society guidelines with a flat improvement along the last 10 years.[26] However, it is also necessary to integrate future ML analysis with other recommendation grades.[27,28]

STUDY LIMITATIONS. First, even though transaortic maximum velocity is the most robust hemodynamic parameter to characterize AS severity, the current definition of severe AS also includes a mean pressure gradient ≥40 mm Hg and an aortic valve area below 1 cm^2.[4] When the model output was redefined on the basis of transaortic peak velocity and mean pressure gradient (severe AS was graded when at least 1 of the following conditions was observed: peak velocity was ≥4 m/s or mean pressure gradient ≥40 mm Hg), only 13 of the 1,638 patients in the training data set and 25 of the 1,533 patients in the external cohort data set would be considered to have severe AS, with this

JACC: CARDIOVASCULAR IMAGING, VOL. 16, NO. 6, 2023
JUNE 2023:733-744

consideration earlier than with the transaortic peak velocity criterion alone. The results from developing new models using these criteria did not yield any improvement with respect to the models presented in the study, both in internal validation or external application (Supplemental Figure 9).

Second, assessment of echocardiographic aortic valve area by continuity equation has well-known theoretical and practical limitations that question its use in patients with mild and moderate aortic valve obstruction and normal transaortic flow.[29,30] Besides, we used a clinical echocardiography cohort to train the model, in which aortic valve area values had not been routinely measured because most of the patients had mild AS, and left ventricular outflow tract diameter measurements were not routinely obtained in these patients. In this setting, introducing aortic valve area as criterion of severe AS implies a selection bias. In any case, when using a dimensionless velocity ratio of 0.25 or less as a definition of severe AS (ie, the ratio of the subvalvular velocity obtained by pulsed wave Doppler imaging to the maximum velocity obtained by continuous wave Doppler), newly developed ML models yielded similar but not improved results compared with the models that were based on the transaortic peak velocity criterion. These findings demonstrated the consistency of the model (Supplemental Figure 10).

The ML model input consists of the echocardiographic data from a single examination, without considering previous echocardiographic studies. The future and prospective incorporation of new databases with serial echocardiographic data will improve the model predictive capacity. Furthermore, valve calcification has been shown to be a good predictor of the progression rate of AS. Although the degree of calcification is qualitatively reflected in most of the echocardiographic reports, the retrospective nature of the analysis does not allow including this variable for semiquantitative grading of valve calcification.

In the developed ML model, misclassification of severe AS was uncommon (3.9%). The interval between misclassification and diagnosis of severe AS was 1 year (**Figure 4**). This interval could be relevant in patients who develop the classic triad of symptoms in the meantime and who should be educated about the possible progression of their valve disease and the need for advancing their echocardiographic examination appointment.

Some patients with AS in our study could have undergone echocardiographic examinations for reasons other than regular follow-up, such as symptom change, preoperative evaluation, other valvular disease, or coexisting diseases. In addition, poor cardiovascular outcomes have been recently related to AS, not always related to disease progression.[31] However, whether prospective echocardiographic monitoring can affect this dismal prognosis remains unknown.

Additional limitations could be model overfitting and complexity because artificial intelligence and ML models can aid the human decision process in routine medical application only if their results are robust and achievable within a reasonable computation time. In this sense, the developed algorithm is fast, and it takes under a second to process the full external cohort data set and provide predictions on a modern PC; conversely, training times are indeed slow, and it can take longer than a day to train the algorithms. Finally, ML applications, although valuable, can have unintended consequences.[32]

CONCLUSIONS

This work showed the efficiency of using ML methodology to provide patient-adaptive follow-up visits on the basis of the quantitative echocardiographic measurements and demographic variables in patients with AS. This methodology has been validated in an external cohort showing the potential to transfer the model among institutions and opening the discussion to use these methodologies as a Level of Evidence: C in current clinical guidelines. However, further prospective validation is required to confirm the results.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This study was supported by the Spanish Cardiovascular Network (CIBERCV) and the Technological Development Projects in Health (DTS19/00098) and by a Research Project in Healthcare (PI21/00369); and was supported by national, public, and competitive resources from the Carlos III Health Institute (Ministry of Science and Innovation, Spain) founded by the European Union European Regional Development Fund. Drs Dorado-Díaz, Sampedro-Gómez, Sánchez-González, Vicente-Palacios, and Sanchez are the inventors of a patent on the method (Aortic Stenosis Echocardiographic Follow-Up Expert System; International Publication No. WO 2020/157212 A1) described in this paper. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Pedro L. Sanchez, Department of Cardiology, Hospital Universitario de Salamanca, Paseo de San Vicente 58, Salamanca 37007, Spain. E-mail: pedrolsanchez@secardiologia.es.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: ML algorithms can be implemented for echocardiographic follow-up of patients with AS because they provide a correct prediction of the progression of valve obstruction.

COMPETENCY IN PATIENT CARE AND PROCEDURAL SKILLS: Echocardiographic follow-up of a patient with AS can be precisely programmed on the basis of the results of an ML algorithm.

TRANSLATIONAL OUTLOOK: The development and implementation of artificial intelligence algorithms in the characterization and follow-up of patients with valve diseases may improve the efficiency of the overall management of these patients. Furthermore, ML could be used to provide expert-level medical assessment for Level of Evidence: C guideline recommendations.

REFERENCES

1. Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. *Lancet*. 2006;368:1005-1011.

2. Otto CM, Prendergast B. Aortic-valve stenosis-from patients at risk to severe valve obstruction. *N Engl J Med*. 2014;371:744-756.

3. Doherty JU, Kort S, Mehran R, Schoenhagen P, Soman P. ACC/AATS/AHA/ASE/ASNC/HRS/SCAI/SCCT/SCMR/STS 2017 appropriate use criteria for multimodality imaging in valvular heart disease: a report of the American College of Cardiology Appropriate Use Criteria Task Force, American Association for Thoracic Surgery, American Heart Association, American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, Society for Cardiovascular Magnetic Resonance, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2017;70:1647-1672.

4. Baumgartner H, Falk V, Bax JJ, et al. 2017 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J*. 2017;38:2739-2791.

5. Chambers JB, Prendergast B, Lung B, et al. Standards defining a 'heart valve centre': ESC Working Group on Valvular Heart Disease and European Association for Cardiothoracic Surgery Viewpoint. *Eur Heart J*. 2017;38:2177-2183.

6. Otto CM, Nishimura RA, Bonow RO, 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.

7. Vahanian A, Beyersdorf F, Praz F, et al. 2021 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J*. 2021;60(4):727-800.

8. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347-1358.

9. Dorado-Díaz PI, Sampedro-Gómez J, Vicente-Palacios V, Sánchez PL. Applications of artificial intelligence in cardiology. the future is already here. *Rev Esp Cardiol (Engl Ed)*. 2019;72:1065-1075.

10. Sengupta PP, Shrestha S, Kagiyama N, et al. A machine-learning framework to identify distinct phenotypes of aortic stenosis severity. *J Am Coll Cardiol Img*. 2021;14:1707-1720.

11. Yang C, Ojha BD, Aranoff ND, Green P, Tavassolian N. Classification of aortic stenosis using conventional machine learning and deep learning methods based on multi-dimensional cardio-mechanical signals. *Sci Rep*. 2020;10:17521.

12. O'Hair DP, Gada H, Sotelo MR, et al. Enhanced detection of heart valve disease using integrated artificial intelligence at scale. *Ann Thorac Surg*. 2022;113:1499-1504.

13. Carrodeguas E, Lacson R, Swanson W, Khorasani R. Use of machine learning to identify follow-up recommendations in radiology reports. *J Am Coll Radiol*. 2019;16:336-343.

14. Dalal S, Hombal V, Weng WH, et al. Determining follow-up imaging study using radiology reports. *J Digit Imaging*. 2020;33:121-130.

15. Ordikhani M, Saniee Abadeh M, Prugger C, Hassannejad R, Mohammadifard N, Sarrafzadegan N. An evolutionary machine learning algorithm for cardiovascular disease risk prediction. *PLoS One*. 2022;17:e0271723.

16. Sengupta PP, Shrestha S, Berthon B, et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): a checklist reviewed by the American College of Cardiology Healthcare Innovation Council. *J Am Coll Cardiol Img*. 2020;13:2017-2035.

17. Gabriel KR. Biplot graphic display of matrices with application to principal component analysis. *Biometrika*. 1971;58:453-467.

18. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.

19. Nielsen D. *Tree Boosting With XGBoost—Why Does XGBoost Win "Every" Machine Learning Competition? Master's thesis. NTNU (Norwegian University of Science and Technology)*. 2016.

20. Rodríguez JD, Pérez A, Lozano JA. Sensitivity analysis of kappa-fold cross validation in prediction error estimation. *IEEE Trans Pattern Anal Mach Intell*. 2010;32:569-575.

21. Pearlman AS, Ryan T, Picard MH, Douglas PS. Evolving trends in the use of echocardiography: a study of Medicare beneficiaries. *J Am Coll Cardiol*. 2007;49:2283-2291.

22. Welch HG, Hayes KJ, Frost C. Repeat testing among Medicare beneficiaries. *Arch Intern Med*. 2012;172:1745-1751.

23. Williams C, Mateescu A, Rees E, et al. Point-of-care echocardiographic screening for left-sided valve heart disease: high yield and affordable cost in an elderly cohort recruited in primary practice. *Echo Res Pract*. 2019;6:71-79.

24. Douglas PS, Garcia MJ, Haines DE, et al. ACCF/ASE/AHA/ASNC/HFSA/HRS/SCAI/SCCM/SCCT/SCMR 2011 appropriate use criteria for echocardiography. a report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, American Society of Echocardiography, American Heart Association, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, Society of Critical Care Medicine, Society of Cardiovascular Computed Tomography, and Society for Cardiovascular Magnetic Resonance Endorsed by the American College of Chest Physicians. *J Am Coll Cardiol*. 2011;57:1126-1166.

25. Steeds RP, Garbi M, Cardim N, et al. EACVI appropriateness criteria for the use of transthoracic echocardiography in adults: a report of literature and current practice review. *Eur Heart J Cardiovasc Imaging*. 2017;18:1191-1204.

26. Fanaroff AC, Califf RM, Windecker S, Smith SC, Lopes RD. Levels of evidence supporting American College of Cardiology/American Heart Association and European Society of Cardiology guidelines, 2008-2018. *JAMA*. 2019;321:1069-1080.

JACC: CARDIOVASCULAR IMAGING, VOL. 16, NO. 6, 2023
JUNE 2023:733-744

27. Tsopra R, Lamy JB, Sedki K. Using preference learning for detecting inconsistencies in clinical practice guidelines: methods and application to antibiotherapy. *Artif Intell Med*. 2018;89:24-33.

28. Feeny AK, Rickard J, Patel D, et al. Machine learning prediction of response to cardiac resynchronization therapy. *Circ Arrhythm Electrophysiol*. 2019;12:e007316.

29. González-Mansilla A, Martinez-Legazpi P, Prieto A, et al. Valve area and the risk of overestimating aortic stenosis. *Heart*. 2019;105:911-919.

30. Pérez Del Villar C, Yotti R, Espinosa M, et al. The functional significance of paradoxical low-gradient aortic valve stenosis: hemodynamic findings during cardiopulmonary exercise testing. *J Am Coll Cardiol Img*. 2017;10:29-39.

31. 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.

32. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318:517-518.

KEY WORDS aortic stenosis, artificial intelligence, echocardiography, follow-up, guidelines, machine learning

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