



# Performance of Prediction Models for Diagnosing Severe Aortic Stenosis Based on Aortic Valve Calcium on Cardiac Computed Tomography: Incorporation of Radiomics and Machine Learning

Nam gyu Kang, MD, Young Joo Suh, MD, PhD, Kyunghwa Han, PhD,  
Young Jin Kim, MD, PhD, Byoung Wook Choi, MD, PhD

All authors: Department of Radiology, Research Institute of Radiological Science, Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

**Objective:** We aimed to develop a prediction model for diagnosing severe aortic stenosis (AS) using computed tomography (CT) radiomics features of aortic valve calcium (AVC) and machine learning (ML) algorithms.

**Materials and Methods:** We retrospectively enrolled 408 patients who underwent cardiac CT between March 2010 and August 2017 and had echocardiographic examinations (240 patients with severe AS on echocardiography [the severe AS group] and 168 patients without severe AS [the non-severe AS group]). Data were divided into a training set (312 patients) and a validation set (96 patients). Using non-contrast-enhanced cardiac CT scans, AVC was segmented, and 128 radiomics features for AVC were extracted. After feature selection was performed with three ML algorithms (least absolute shrinkage and selection operator [LASSO], random forests [RFs], and eXtreme Gradient Boosting [XGBoost]), model classifiers for diagnosing severe AS on echocardiography were developed in combination with three different model classifier methods (logistic regression, RF, and XGBoost). The performance (c-index) of each radiomics prediction model was compared with predictions based on AVC volume and score.

**Results:** The radiomics scores derived from LASSO were significantly different between the severe AS and non-severe AS groups in the validation set (median, 1.563 vs. 0.197, respectively,  $p < 0.001$ ). A radiomics prediction model based on feature selection by LASSO + model classifier by XGBoost showed the highest c-index of 0.921 (95% confidence interval [CI], 0.869–0.973) in the validation set. Compared to prediction models based on AVC volume and score (c-indexes of 0.894 [95% CI, 0.815–0.948] and 0.899 [95% CI, 0.820–0.951], respectively), eight and three of the nine radiomics prediction models showed higher discrimination abilities for severe AS. However, the differences were not statistically significant ( $p > 0.05$  for all).

**Conclusion:** Models based on the radiomics features of AVC and ML algorithms may perform well for diagnosing severe AS, but the added value compared to AVC volume and score should be investigated further.

**Keywords:** Aortic valve calcium; Aortic stenosis; Computed tomography; Radiomics; Machine learning

## INTRODUCTION

Aortic stenosis (AS) is the most common valvular

heart disease in developed countries. In severe AS, the timing of surgery is vital for preventing subsequent cardiovascular events. Although AS severity is assessed by

**Received:** February 6, 2020 **Revised:** May 12, 2020 **Accepted:** June 4, 2020

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2018R1C1B6007251) and a faculty research grant of Yonsei University College of Medicine (6-2018-0041).

**Corresponding author:** Young Joo Suh, MD, PhD, Department of Radiology, Research Institute of Radiological Science, Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.

• E-mail: rongzusuh@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

echocardiography as a standard reference, these assessments can often be inconclusive (1, 2). Most cases that are inconclusive on echocardiography are low-gradient AS, and measuring the aortic valve calcium (AVC) score on cardiac computed tomography (CT) is helpful. The AVC burden on CT is positively associated with AS severity, and AVC score on CT predicts outcomes better than clinical assessments and Doppler echocardiography results (3-5). In this context, the 2017 European Society of Cardiology/European Association for Cardio-Thoracic Surgery guidelines recommend CT AVC scoring in cases of low-flow, low-gradient AS with preserved ejection (6).

Although AVC plays a key role in AS progression, it is not the only factor that affects the degree of AS; some patients with low AVC are diagnosed with hemodynamically severe AS (7). AS severity is influenced by other factors, including AVC attenuation, shape, symmetry, or distribution. As shown in a previous study, not only the degree of AVC but also the localization of AVC within the valve is associated with AS severity (8). The term “radiomics” refers to the high-throughput extraction of high-dimensional quantitative information from a medical image. Although radiomics has been studied primarily in oncology, the application of radiomics to cardiovascular imaging has been growing (9-13). Because radiomics features provide quantitative information about a region-of-interest (ROI), such as volume, shape, texture, and high-order features, analysis of these features may yield data that are predictive of AS severity. Additionally, because radiomics data are high-dimensional, appropriate methods for feature selection and model classifiers are required, and the application of machine learning is expected to be helpful (14, 15). Therefore, our study aimed to develop a prediction model for diagnosing severe AS using the radiomics features of AVC on CT and machine learning algorithms.

## MATERIALS AND METHODS

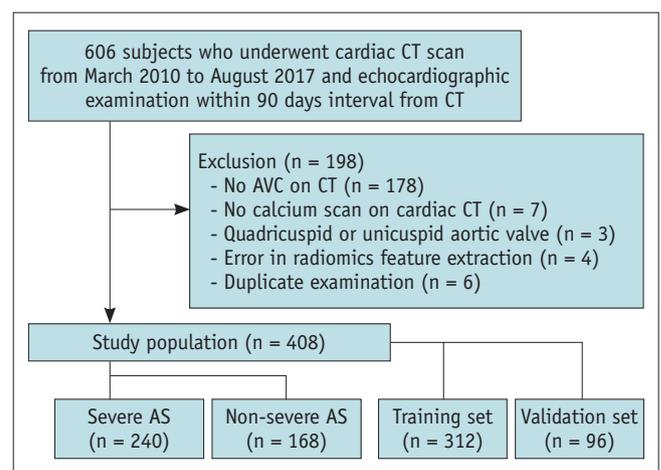
### Study Population

Our Institutional Review Board approved this study, and the requirement for informed consent was waived (IRB number: 4-2019-0006). We retrospectively identified 606 patients with valvular heart disease who underwent cardiac CT between March 2010 and August 2017 and had echocardiographic examinations within 90 days of the CT scan. Patients were excluded based on the following criteria: 1) patients with no AVC on cardiac CT because radiomics

features could not be extracted from the segmented AVC (n = 178), 2) patients who did not undergo coronary calcium scans during cardiac CT evaluation (n = 7), 3) patients who had a quadricuspid or unicuspid aortic valve (AV) (n = 3), or 4) patients with errors in radiomics feature extraction (n = 4). For the 6 patients who had two cardiac CT scans during the study period that were performed within 90 days of the echocardiography, only the most recent CT was included. A total of 408 patients were enrolled in this study (208 men; mean age, 66.0 ± 10.9 years) (Fig. 1). Data were divided into a training set (312 patients who underwent CT between March 2010 and December 2016) and a validation set (96 patients who underwent CT between January 2017 and August 2017) with reference to a specific time point. The use of the term “validation” was similar to that used in the medicine- and health-related literature, indicating a process used to verify model performance (16).

### CT Image Acquisition and Reconstruction

All cardiac CT scans were performed using a second-generation, dual-source scanner (SOMATOM Definition Flash, Siemens Healthineers) or a 256-detector-row CT scanner (Revolution CT, GE Healthcare). Non-contrast-enhanced calcium scans were performed using a prospective electrocardiogram-gated sequential scan mode with a tube voltage of 120 kV and a tube current-time product of 50 mAs. The images were reconstructed with a slice thickness of 2.5–3 mm and an increment of 2.5–3 mm with a medium-sharp kernel and filtered back projection (Supplementary Materials 1).



**Fig. 1. Flowchart of the study population.** AS = aortic stenosis, AVC = aortic valve calcium, CT = computed tomography

### Segmentation of AVC and Radiomics Feature Extraction

From axial images of calcium scoring CT, two radiologists (a fourth-year senior radiology resident, and a board-certified radiologist with 7 years of experience in cardiothoracic imaging), who were blinded to the echocardiographic results and the other observer's segmentation results, independently segmented AVC using a commercial software (AVIEW Research, Coreline Soft Inc.) (Supplementary Fig. 1). One observer repeated the AVC segmentation for 100 randomly selected cases with a time interval of at least 2 months from the first segmentation to assess intraobserver variability. AVC was defined as a region  $\geq 1 \text{ mm}^3$  with a density of  $\geq 130$  Hounsfield units at the AV leaflets and annulus. Observers segmented AVC by carefully including ROIs in the AV leaflet and annulus and excluding calcium in the adjacent sinus of Valsalva, left ventricular outflow tract, or mitral annulus, and image noise or beam-hardening artifact was excluded (17). From the segmented ROI, the AVC volume was measured, and an AVC score was calculated using the Agatston method (18).

From the segmented ROI of AVC, 128 radiomics features were extracted using the AVIEW software. The radiomics features were categorized as follows: 1) size and shape features, 2) first-order features based on the histogram, 3) gray level co-occurrence matrix features, 4) gray level run length matrix features, 5) gray level size zone matrix features, 6) gray level dependence matrix features, 7) neighborhood gray-tone difference matrix features, 8) moment features, 9) gradient features, and 10) fractal features. A list of the specific features contained in each category is described in Supplementary Table 1.

### Assessment of the Severity of AS by Echocardiography

All patients underwent a comprehensive transthoracic or transesophageal echocardiography examination (interval from cardiac CT, median 1 day; interquartile range [IQR], 0–9.5 days). The grading of AS severity was evaluated based on peak velocity, mean gradient, and calculation of the AV area using the continuity equation, as recommended by the current guideline: mild AS, mean gradient  $< 20 \text{ mm Hg}$  or peak aortic jet velocity of 2.0–2.9 m/s; moderate AS, mean gradient of 20–39 mm Hg or peak aortic jet velocity of 3.0–3.9 m/s; severe AS, mean gradient  $\geq 40 \text{ mm Hg}$  or peak aortic jet velocity  $\geq 4.0 \text{ m/s}$  and AV area  $\leq 1 \text{ cm}^2$  (indexed AV area by body surface area,  $< 0.6 \text{ cm}^2/\text{m}^2$ ); very severe AS, mean gradient  $\geq 60 \text{ mm Hg}$  or peak aortic jet velocity  $\geq 5.0 \text{ m/s}$  (19).

### Radiomics Feature Selection

Using the data from the training set, radiomics feature selection was performed with three different machine learning methods: least assembly shrinkage and selection operator (LASSO), random forests (RFs), and eXtreme Gradient Boosting (XGBoost). LASSO was performed with fivefold cross-validation to overcome the overfitting problem (14), and features showing nonzero coefficients using LASSO-logistic regression were selected. A radiomics score was calculated for each case via a linear combination of selected features that were weighted by their respective coefficient calculated by the LASSO-logistic regression model. In RF, feature selection was performed with variable importance (VIMP) (20). In XGBoost, feature selection was performed according to the gain, which is the relative contribution of the corresponding feature to the model calculated by considering the contribution of each feature to each tree in the model (21). RF and XGBoost are bootstrap and boosting-based methods, respectively; both methods are used to diminish the overfitting problem. Feature selection was performed using Python (version 3.6.7), Scikit-learn (version 0.20.1), and XGBoost (version 0.82). The feature selection method, "SelectFromModel," was used with RF and XGBoost.

### Classifier Model for Diagnosing Severe AS

For data analysis, we divided patients into two groups according to the AS severity on echocardiography: the severe AS group (severe or very severe AS) and the non-severe AS group (no, mild, or moderate AS). Using the volume of AVC and the AVC score, prediction models for diagnosing severe AS were constructed based on logistic regression. Using the radiomics features selected by each of the machine learning methods, classifier models for diagnosing severe AS on echocardiography were developed with three different classifier methods: traditional logistic regression based on radiomics features and two machine learning methods (RF and XGBoost). Therefore, a total of nine prediction models were developed by combining the three radiomics feature selection methods and the three classifier methods. All classifier models were developed with fivefold cross-validation.

### Validation of the Prediction Models

The prediction models built with the training set were applied to the validation set. The performance of each model was evaluated based on the c-index, which

represents the discrimination ability of each model using the validation set.

### Statistical Analyses

All statistical analyses except feature selection and classifier modeling were performed using a statistical software package (SPSS version 26.0, IBM Corp.). Categorical variables are expressed as the number with percentage and compared using the chi-squared test. Continuous variables are expressed as the mean with standard deviation or as the median with IQR according to the normality of the data and compared using the Student's *t* test or the Mann-Whitney U test. Intra- and inter-observer agreement of the segmented AVC volume was assessed using the intraclass correlation coefficient. The performance of each model in diagnosing severe AS based on radiomics features was compared with models based on the AVC volume and AVC score using Delong's method (22). For subgroup analysis, we divided the patients of the validation set into two groups according to the AVC volume (low and high) and AVC score (low and high). The same prediction

models used for the entire validation set were used for the subgroup analysis. The best cutoff value for AVC volume for predicting severe AS was calculated based on the Youden index from the training set data (23). For the AVC score, the known threshold for severe AS was used as  $\geq 2000$  for men and  $\geq 1200$  for women (4, 6). A probability value of less than 0.05 was considered statistically significant.

## RESULTS

### Patients

Table 1 lists the demographic statistics of the 312 patients in the training set and 96 patients in the validation set. Clinical characteristics of patients in the severe AS and non-severe AS groups showed no significant differences in training and validation sets, except for age in the training set. No significant difference in patient characteristics was observed between training and validation sets. AVC volume and score were significantly different between the severe AS and non-severe AS groups in training and validation sets (Supplementary Table 2).

**Table 1. Patient Characteristics**

	Training Set (n = 312)			<i>P</i> *	Validation Set (n = 96)			<i>P</i> *	<i>P</i> between Training and Validation Set
	Total	Severe AS Group (n = 180)	Non-Severe AS Group (n = 132)		Total	Severe AS Group (n = 60)	Non-Severe AS Group (n = 36)		
Mean age (years)	65.9 ± 11.3	67.3 ± 10.7	63.91 ± 12.0	0.040	66.4 ± 9.7	66.1 ± 9.4	66.9 ± 10.2	0.574	0.050
Male sex	164 (52.5)	93 (51.7)	71 (53.8)	0.732	52 (54.1)	35 (58.3)	17 (47.2)	0.398	0.816
Body mass index (kg/m <sup>2</sup> )	23.8 ± 3.2	24.1 ± 3.2	23.4 ± 3.3	0.055	23.5 ± 2.9	23.6 ± 3.2	23.4 ± 2.6	0.751	0.436
Body surface area (m <sup>2</sup> )	1.66 ± 0.18	1.66 ± 0.18	1.67 ± 0.19	0.830	1.65 ± 0.18	1.66 ± 0.19	1.63 ± 0.16	0.407	0.511
Hypertension	199 (63.8)	121 (67.2)	78 (59.1)	0.141	59 (61.5)	34 (56.7)	25 (69.4)	0.215	0.680
Diabetes mellitus	62 (19.9)	39 (21.7)	23 (17.4)	0.354	24 (25.0)	19 (31.7)	5 (13.9)	0.053	0.282
Dyslipidemia	45 (14.4)	24 (13.3)	21 (15.9)	0.523	20 (20.8)	14 (23.3)	6 (16.7)	0.439	0.134
AS severity				< 0.001				< 0.001	0.603
Normal	64 (20.5)	0 (0.0)	64 (48.5)		14 (14.6)	0 (0.0)	14 (38.9)		
Mild	17 (5.4)	0 (0.0)	17 (12.9)		7 (7.3)	0 (0.0)	7 (19.4)		
Moderate	51 (16.3)	0 (0.0)	51 (38.6)		15 (15.6)	0 (0.0)	15 (41.7)		
Severe	160 (51.3)	160 (88.9)	0 (0.0)		51 (53.1)	51 (85.0)	0 (0.0)		
Very severe	20 (6.4)	20 (11.1)	0 (0.0)		9 (9.4)	9 (15.0)	0 (0.0)		
Valve type				0.002				0.013	0.807
Bicuspid	109 (34.9)	76 (42.2)	33 (25.0)		32 (33.3)	26 (43.3)	6 (16.7)		
Tricuspid	203 (65.1)	104 (57.8)	99 (75.0)		64 (66.7)	34 (56.7)	30 (83.3)		

Unless stated otherwise, values are presented as n (%). \*Indicates *p* value for comparison between severe AS group and non-severe AS group. AS = aortic stenosis

**Radiomics Feature Extraction**

Inter-observer and intra-observer agreement for AVC volume segmentation was excellent showing intraclass correlation coefficients of 0.992 (95% confidence interval [CI], 0.978–0.996) and 0.989 (95% CI, 0.977–0.994), respectively. A majority of the 128 extracted radiomics features showed statistically significant differences between the severe AS and non-severe AS groups in the training set ( $p < 0.05$ ), except six features (Supplementary Table 2).

**Radiomics Feature Selection**

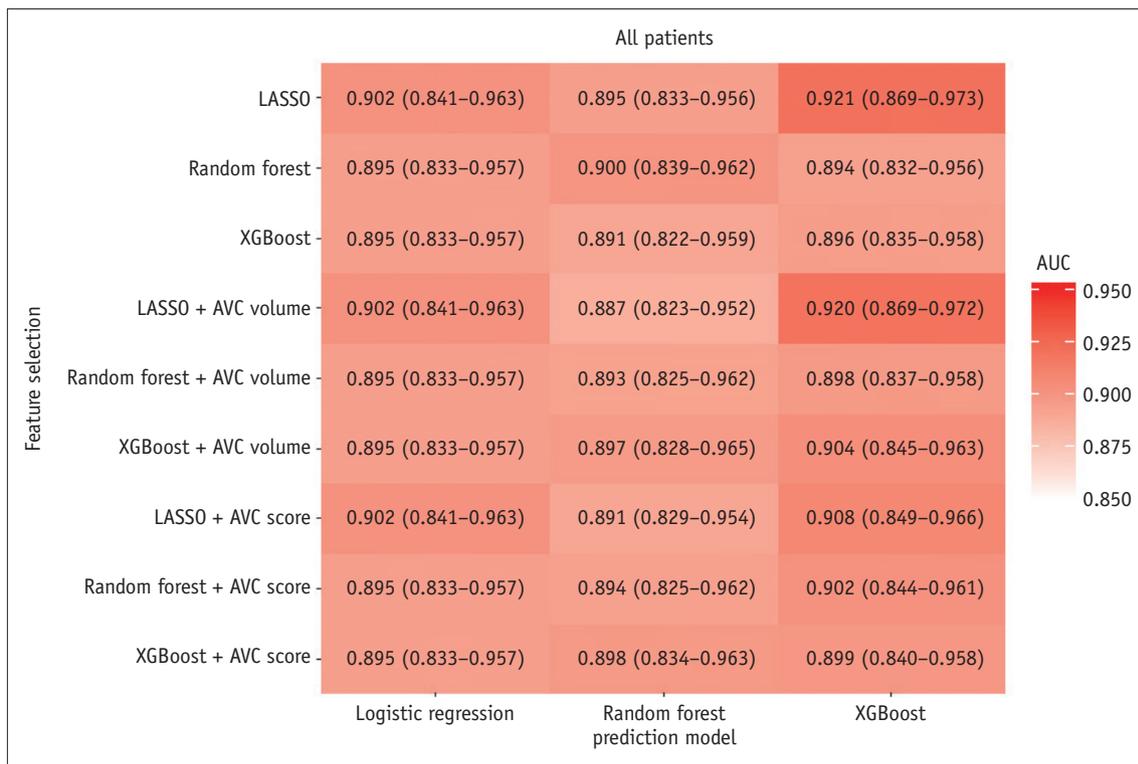
Using LASSO, 19 features were selected in the training set, and the radiomics score was calculated based on these 19 features (Supplementary Materials 2). Using RF, 82 radiomics features were selected with a VIMP threshold higher than 0.00411 (Supplementary Table 3), and using XGBoost, 79 features were selected with a gain score threshold higher than 0.00261 (Supplementary Table 4).

The radiomics score derived from LASSO was significantly different between the non-severe AS and severe AS groups in the validation set (median 0.197, IQR 0.0572–0.543 vs. median 1.563, IQR 0.681–2.879, respectively,  $p < 0.001$ ).

**Comparison of Model Performance for Diagnosing Severe AS in the Validation Set**

Prediction models based on AVC volume alone and AVC score alone had c-indexes of 0.894 (95% CI, 0.815–0.948) and 0.899 (95% CI, 0.820–0.951), respectively. The c-indexes of various combinations of prediction models for diagnosing severe AS with the validation set are presented in Figure 2. The radiomics prediction model based on the combination of feature selection by LASSO and model classifier by XGBoost had the highest c-index of 0.921 (95% CI, 0.869–0.973) in the validation set, whereas the lowest c-index of 0.891 (95% CI, 0.822–0.959) was obtained with the combination of feature selection by XGBoost and model classifier by RF.

Compared with the prediction model based on AVC volume (c-index = 0.894), eight of the nine radiomics prediction models showed higher predictability. However, no statistically significant difference was observed in any comparison (Supplementary Table 5). Compared with the prediction model based on AVC score (c-index = 0.899), three of the nine prediction models showed higher predictability, although the differences were not statistically significant.

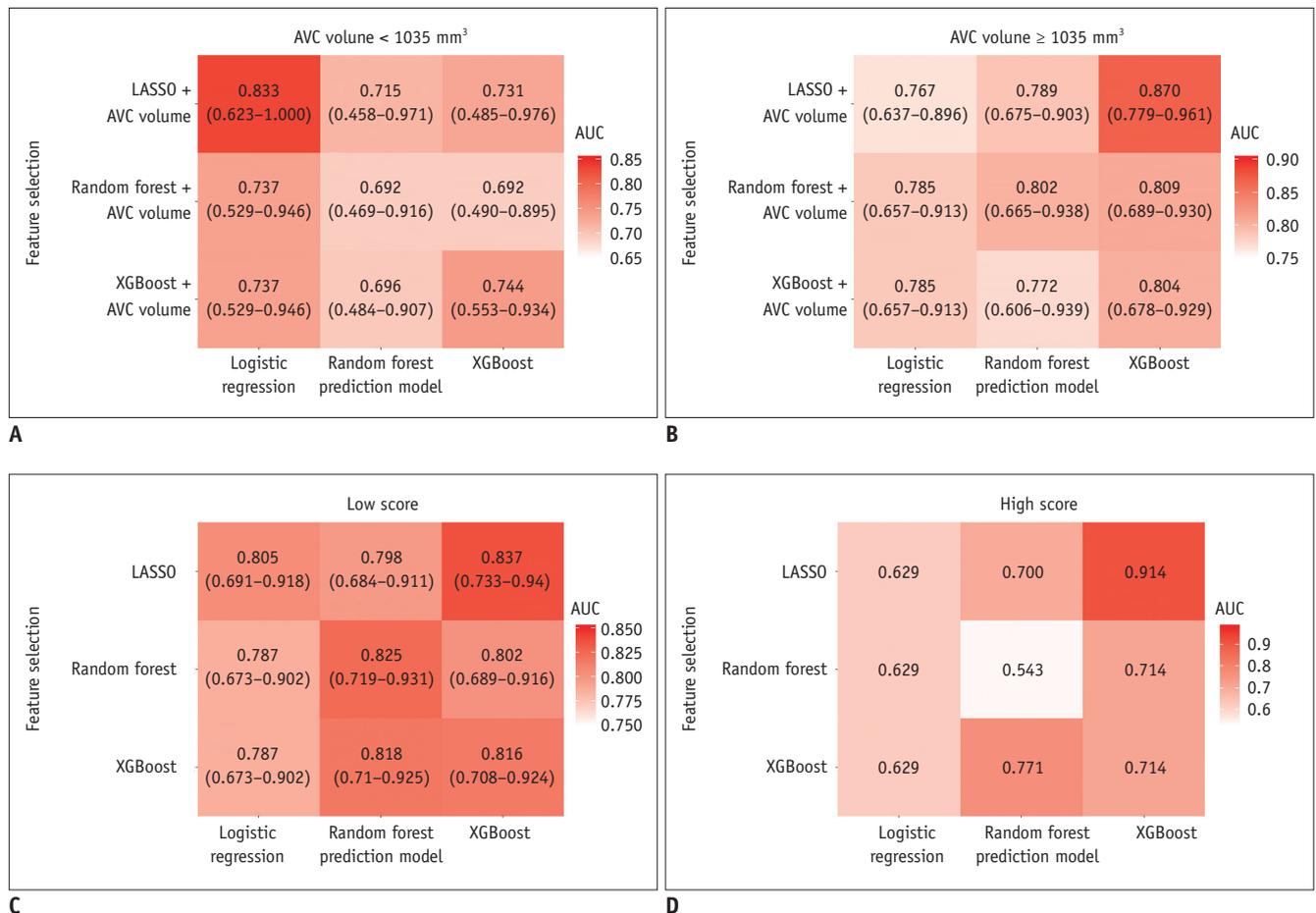


**Fig. 2. Heatmap showing the discrimination ability of prediction models for severe AS according to various combinations of feature selection and model classifier methods.** For example, the model based on the feature selection method “random forests” and model classifier method “XGBoost” shows an AUC of 0.894 (95% CI, 0.832–0.956), and another model based on the feature selection method “XGBoost” and model classifier method “random forests” shows an AUC of 0.891 (95% CI, 0.822–0.959). AUC = area under the curve, CI = confidence interval, LASSO = least assembly shrinkage and selection operator, XGBoost = eXtreme Gradient Boosting

**Subgroup Analysis**

The optimal cutoff value of AVC volume for severe AS from the training set was 1035 mm<sup>3</sup>. For the subgroup with AVC volumes < 1035 mm<sup>3</sup> in the validation set (n = 32), the prediction model based on AVC volume alone had a c-index of 0.787 (95% CI, 0.657–0.916). The radiomics prediction model for diagnosing severe AS with the highest c-index of 0.833 was based on the combination of feature selection by LASSO and model classifier by logistic regression (95% CI, 0.623–1.000) (Fig. 3). In the subgroup with AVC volume ≥ 1035 mm<sup>3</sup> in the validation set (n = 64), the prediction model based on AVC volume alone had a c-index of 0.798 (95% CI, 0.570–1.000). The radiomics prediction model with the highest c-index of 0.870 was based on the combination of feature selection by LASSO and model classifier by XGBoost (95% CI, 0.779–0.961). However, no statistically significant difference was observed in any comparison with the model based on AVC volume (Supplementary Table 6).

When a predefined threshold was used, prediction using AVC score had a c-index of 0.778 (95% CI, 0.681–0.856), sensitivity of 58.3%, and specificity of 97.2% for diagnosing severe AS in the validation set. For the subgroup with AVC scores lower than the threshold in the validation set (n = 36), the prediction model based on AVC score alone had a c-index of 0.804 (95% CI, 0.694–0.915). The radiomics prediction model for diagnosing severe AS with the highest c-index of 0.837 was based on the combination of feature selection by LASSO and model classifier by XGBoost (95% CI, 0.733–0.940) (Fig. 3). For the subgroup with AVC scores greater than or equal to the threshold in the validation set (n = 60), the prediction model based on AVC score alone had a c-index of 0.647 (95% CI could not be calculated because only one case was assigned as having non-severe AS in this subgroup). The highest c-index of 0.914 was obtained with the combination of feature selection by LASSO and model classifier by XGBoost.



**Fig. 3. Performance of models in subgroup according to AVC volume and AVC score.** **A.** Subgroup with low AVC volume (< 1035 mm<sup>3</sup>, n = 32). **B.** Subgroup with high AVC volume (≥ 1035 mm<sup>3</sup>, n = 64). **C.** Subgroup with low AVC score (< 2000 for men and < 1200 for women, n = 36). **D.** Subgroup with high AVC score (≥ 2000 for men and ≥ 1200 for women, n = 60). 95% CI could not be provided due to small number of patients with non-severe AS (n = 1) in this subgroup.

However, no statistically significant difference was observed in any comparison with the model based on AVC score (Supplementary Table 6).

## DISCUSSION

Our study demonstrates that prediction models based on a combination of AVC radiomics features derived from non-contrast-enhanced cardiac CT and machine learning algorithms diagnose severe AS better than models based on AVC volume or AVC score alone, although the differences are not statistically significant. The prediction models based on AVC volume and AVC score alone show discrimination abilities with c-indexes of 0.894 and 0.899, respectively, but the models constructed by combining radiomics features and machine learning have c-indexes as high as 0.921. In subgroups with low AVC volume ( $< 1035 \text{ mm}^3$ ) or low AVC score, the radiomics prediction model yields c-indexes of 0.833 and 0.837, respectively, for diagnosing severe AS.

AVC is closely associated with AS severity, and AVC burden measured by CT helps diagnose severe AS when Doppler parameters are inconclusive, specifically in patients with suspected low-flow, low-gradient AS. For the quantification of AVC, Agatston scoring was the most frequently used method, exhibiting good diagnostic performance for determining severe AS, having c-indexes of 0.81–0.91, sensitivities of 80–89.7%, and specificities of 77–89% (3, 4, 24–26). However, in our study, using AVC score to diagnose severe AS had a c-index of 0.778, sensitivity of 58.3%, and specificity of 97.2% in the validation set. These values of c-index and sensitivity are lower than those reported in previous studies.

Previous studies of AVC score on CT used different cutoff values for the calculated scores (3, 4, 24–26). Some studies suggested sex-specific cutoff values ( $\geq 1200$  Agatston score for women and  $\geq 2000$  Agatston score for men) for diagnosing severe AS, and the current guideline accepted these cutoff values (4, 6, 26). Despite the high performance of AVC burden for diagnosing severe AS reported in previous studies, in our study population, the AVC burden exhibited a notably lower diagnostic performance. We assume that this may be due to inter-study differences in patient characteristics, such as race and valve morphology type (27, 28). Compared with previous studies conducted in Western countries, our study population comprised East Asians, who have smaller body sizes and smaller AV cusps. Moreover, the proportion of patients with bicuspid valves

is higher in our population (approximately one-third) than that in other populations. These population differences may have contributed to the lower diagnostic performance of AVC burden in our study because different annular and leaflet geometries, as well as differences in the distribution and burden of AVC, can be expected among patients with bicuspid AVs.

The application of radiomics in cardiac imaging has been limited to specific topics, such as the identification of vulnerable coronary plaques or differentiation of cardiomyopathy (9–11). To the best of our knowledge, this study is the first to apply radiomics to the assessment of AVC. Because AS severity is not always proportional to the AVC burden, we expected that characteristics other than AVC burden, such as morphology, texture, and distribution, might contribute to AS severity. A previous study suggested that the tomographic pattern of AVC could be an independent determinant of AS severity because the CT attenuation ratio of the center to the periphery of the AV was higher in severe AS than in moderate AS (8). Accordingly, our study demonstrates promising results. Even in cases with less severe AVC burden, the prediction model based on radiomics and machine learning algorithms discriminated severe AS with high diagnostic performance.

Because radiomics contain high-dimensional data, proper strategies for feature selection and model classifiers are required, and machine learning algorithms can be effective for these purposes. The combination of feature selection by LASSO and model classifier by XGBoost demonstrated the highest performance for diagnosing severe AS in the validation set and most subgroups. LASSO is an excellent method for feature selection because it retains the useful features of subset selection and ridge regression. It is suitable for analyzing large sets of radiomics features with relatively small sample sizes. RF is an efficient decision tree-based method for variable selection and classification, remains robust when data contain noise and outliers, and can handle high-dimensional spaces rapidly. However, RF is more subjective to an overfitting problem and is considered to be better suited for handling “tall” data, which has larger sample sizes than “wide” data. XGBoost is an ensemble algorithm of decision trees (21). The ensemble works by combining a set of weaker machine learning algorithms to obtain an overall machine learning algorithm that is more robust. The main difference between XGboost and RF is the manner of sampling. RF is based on uniform sampling with return, whereas XGboost assigns higher weights to

wrongly predicted samples in the current weaker learner. Subsequently, these samples will be given more attention when training the next weaker learner. Additionally, XGboost adds regularization to avoid overfitting. Therefore, XGboost is more complicated than RF, and it often outperforms RF. Although standard machine learning methods for feature selection and model classifiers have not been established, we speculated that LASSO would be efficient for feature selection and XGboost would be an appropriate model classifier for radiomics analysis in our study due to the reasons mentioned above. In addition to predicting AS severity, AVC burden has clinical implications for overall survival and post-procedural outcomes after transcatheter AV implantation (5, 29-33). Comprehensive morphological information based on radiomics analysis with efficient machine learning algorithms has the potential to predict not only severe AS but also subsequent clinical outcomes.

Our study has several limitations. First, it was a retrospective study that recruited patients from a single institution, and external validation was not performed. Instead, internal validation was performed by dividing the dataset into a training set and a validation set. Further study is required to validate our results externally and generalize the utility of the radiomics analysis of AVC. Second, the reproducibility of radiomics features was not fully considered, specifically regarding the effects of the CT scanner, image acquisition, and reconstruction parameters on the extracted radiomics features and prediction models (34, 35). Application of deep learning algorithm may help solve the problem of reproducibility in radiomics by reducing variability in scan protocol, for example, conversion of reconstruction kernel or slice thickness (36-38). Third, AVC is currently recommended for the diagnosis of low-flow, low-gradient AS (6). However, we did not analyze the value of AVC radiomics features for diagnosing this specific subtype of AS due to the small number of patients with this condition. Finally, information such as AVC attenuation and clinical variables (e.g., sex and valve morphology) were not considered when constructing the prediction model. Combining these data would improve the performance of the model and should be investigated in future studies.

In conclusion, prediction models based on the radiomics features of AVC and machine learning algorithms may perform well for diagnosing severe AS, but the added value compared to AVC volume and score should be investigated further.

## Supplementary Materials

The Data Supplement is available with this article at <https://doi.org/10.3348/kjr.2020.0099>.

## Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

## ORCID iDs

Nam gyu Kang

<https://orcid.org/0000-0002-7992-7858>

Young Joo Suh

<https://orcid.org/0000-0002-2078-5832>

Kyunghwa Han

<https://orcid.org/0000-0002-5687-7237>

Young Jin Kim

<https://orcid.org/0000-0002-6235-6550>

Byoung Wook Choi

<https://orcid.org/0000-0002-8873-5444>

## REFERENCES

1. Clavel MA, Burwash IG, Pibarot P. Cardiac imaging for assessing low-gradient severe aortic stenosis. *JACC Cardiovasc Imaging* 2017;10:185-202
2. Berthelot-Richer M, Pibarot P, Capoulade R, Dumesnil JG, Dahou A, Thebault C, et al. Discordant grading of aortic stenosis severity: echocardiographic predictors of survival benefit associated with aortic valve replacement. *JACC Cardiovasc Imaging* 2016;9:797-805
3. Koos R, Mahnken AH, Sinha AM, Wildberger JE, Hoffmann R, Kühl HP. Aortic valve calcification as a marker for aortic stenosis severity: assessment on 16-MDCT. *AJR Am J Roentgenol* 2004;183:1813-1818
4. Pawade T, Clavel MA, Tribouilloy C, Dreyfus J, Mathieu T, Tastet L, et al. Computed tomography aortic valve calcium scoring in patients with aortic stenosis. *Circ Cardiovasc Imaging* 2018;11:e007146
5. Clavel MA, Pibarot P, Messika-Zeitoun D, Capoulade R, Malouf J, Aggarwal S, et al. Impact of aortic valve calcification, as measured by MDCT, on survival in patients with aortic stenosis: results of an international registry study. *J Am Coll Cardiol* 2014;64:1202-1213
6. Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ, et al. 2017 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur Heart J* 2017;38:2739-2791
7. Abramowitz Y, Jilaihawi H, Pibarot P, Chakravarty T, Kashif M, Kazuno Y, et al. Severe aortic stenosis with low aortic valve calcification: characteristics and outcome following

- transcatheter aortic valve implantation. *Eur Heart J Cardiovasc Imaging* 2017;18:639-647
8. de Santis A, Tarasoutchi F, Araujo Filho JAB, Vieira MC, Nomura CH, Katz M, et al. Topographic pattern of valve calcification: a new determinant of disease severity in aortic valve stenosis. *JACC Cardiovasc Imaging* 2018;11:1032-1035
  9. Kolossváry M, Park J, Bang JI, Zhang J, Lee JM, Paeng JC, et al. Identification of invasive and radionuclide imaging markers of coronary plaque vulnerability using radiomic analysis of coronary computed tomography angiography. *Eur Heart J-Cardiovasc Imaging* 2019;20:1250-1258
  10. Kolossváry M, Karády J, Szilveszter B, Kitslaar P, Hoffmann U, Merkely B, et al. Radiomic features are superior to conventional quantitative computed tomographic metrics to identify coronary plaques with napkin-ring sign. *Circ Cardiovasc Imaging* 2017;10:e006843
  11. Neisius U, El-Rewaidy H, Nakamori S, Rodriguez J, Manning WJ, Nezafat R. Radiomic analysis of myocardial native T1 imaging discriminates between hypertensive heart disease and hypertrophic cardiomyopathy. *JACC Cardiovasc Imaging* 2019;12:1946-1954
  12. Baessler B, Luecke C, Lurz J, Klingel K, von Roeder M, de Waha S, et al. Cardiac MRI texture analysis of T1 and T2 maps in patients with infarctlike acute myocarditis. *Radiology* 2018;289:357-365
  13. Nam K, Suh YJ, Han K, Park SJ, Kim YJ, Choi BW. Value of computed tomography radiomic features for differentiation of periprosthetic mass in patients with suspected prosthetic valve obstruction. *Circ Cardiovasc Imaging* 2019;12:e009496
  14. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives. *Korean J Radiol* 2019;20:1124-1137
  15. Park YW, Choi YS, Ahn SS, Chang JH, Kim SH, Lee SK. Radiomics MRI phenotyping with machine learning to predict the grade of lower-grade gliomas: a study focused on nonenhancing tumors. *Korean J Radiol* 2019;20:1381-1389
  16. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology formMedical diagnosis and prediction. *Radiology* 2018;286:800-809
  17. Eberhard M, Hinzpeter R, Polacin M, Morsbach F, Maisano F, Nietlispach F, et al. Reproducibility of aortic valve calcification scoring with computed tomography - An interplatform analysis. *J Cardiovasc Comput Tomogr* 2019;13:92-98
  18. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M Jr, Detrano R. Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol* 1990;15:827-832
  19. Nishimura RA, Otto CM, Bonow RO, Carabello BA, Erwin JP 3rd, Guyton RA, et al. 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 Practice Guidelines. *J Thorac Cardiovasc Surg* 2014;148:e1-e132
  20. Breiman L. Random forests. *Machine learning* 2001;45:5-32
  21. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*; 2016 Aug; San Francisco, CA, USA: Association for Computing Machinery; p. 785-794
  22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-845
  23. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J* 2005;47:458-472
  24. Shimizu K, Yamamoto M, Koyama Y, Kodama A, Sato H, Kano S, et al. Usefulness of routine aortic valve calcium score measurement for risk stratification of aortic stenosis and coronary artery disease in patients scheduled cardiac multislice computed tomography. *Int J Cardiol Heart Vasc* 2015;9:95-99
  25. Cueff C, Serfaty JM, Cimadevilla C, Laissy JP, Himbert D, Tubach F, et al. Measurement of aortic valve calcification using multislice computed tomography: correlation with haemodynamic severity of aortic stenosis and clinical implication for patients with low ejection fraction. *Heart* 2011;97:721-726
  26. Clavel MA, Messika-Zeitoun D, Pibarot P, Aggarwal SR, Malouf J, Araoz PA, et al. The complex nature of discordant severe calcified aortic valve disease grading: new insights from combined Doppler echocardiographic and computed tomographic study. *J Am Coll Cardiol* 2013;62:2329-2338
  27. Chandra S, Lang RM, Nicolarsen J, Gayat E, Spencer KT, Mor-Avi V, et al. Bicuspid aortic valve: inter-racial difference in frequency and aortic dimensions. *JACC Cardiovasc Imaging* 2012;5:981-989
  28. Kong WKF, Regeer MV, Poh KK, Yip JW, van Rosendaal PJ, Yeo TC, et al. Inter-ethnic differences in valve morphology, valvular dysfunction, and aortopathy between Asian and European patients with bicuspid aortic valve. *Eur Heart J* 2018;39:1308-1313
  29. Azzalini L, Ghoshhajra BB, Elmariah S, Passeri JJ, Inglessis I, Palacios IF, et al. The aortic valve calcium nodule score (AVCNS) independently predicts paravalvular regurgitation after transcatheter aortic valve replacement (TAVR). *J Cardiovasc Comput Tomogr* 2014;8:131-140
  30. Khalique OK, Hahn RT, Gada H, Nazif TM, Vahl TP, George I, et al. Quantity and location of aortic valve complex calcification predicts severity and location of paravalvular regurgitation and frequency of post-dilation after balloon-expandable transcatheter aortic valve replacement. *JACC Cardiovasc Interv* 2014;7:885-894
  31. Ewe SH, Ng AC, Schuijff JD, van der Kley F, Colli A, Palmen M, et al. Location and severity of aortic valve calcium and implications for aortic regurgitation after transcatheter aortic

- valve implantation. *Am J Cardiol* 2011;108:1470-1477
32. Jilaihawi H, Makkar RR, Kashif M, Okuyama K, Chakravarty T, Shiota T, et al. A revised methodology for aortic-valvar complex calcium quantification for transcatheter aortic valve implantation. *Eur Heart J Cardiovasc Imaging* 2014;15:1324-1332
  33. Hansson NC, Norgaard BL, Barbanti M, Nielsen NE, Yang TH, Tamburino C, et al. The impact of calcium volume and distribution in aortic root injury related to balloon-expandable transcatheter aortic valve replacement. *J Cardiovasc Comput Tomogr* 2015;9:382-392
  34. Berenguer R, Pastor-Juan MDR, Canales-Vazquez J, Castro-Garcia M, Villas MV, Mansilla Legorburo F, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology* 2018;288:407-415
  35. Sung P, Lee JM, Joo I, Lee S, Kim TH, Ganeshan B. Evaluation of the impact of iterative reconstruction algorithms on computed tomography texture features of the liver parenchyma using the filtration-histogram method. *Korean J Radiol* 2019;20:558-568
  36. Lee SM, Lee JG, Lee G, Choe J, Do KH, Kim N, et al. CT image conversion among different reconstruction kernels without a sinogram by using a convolutional neural network. *Korean J Radiol* 2019;20:295-303
  37. Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al. Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology* 2019;292:365-373
  38. Park S, Lee SM, Do KH, Lee JG, Bae W, Park H, et al. Deep learning algorithm for reducing CT slice thickness: effect on reproducibility of radiomic features in lung cancer. *Korean J Radiol* 2019;20:1431-1440