

## ORIGINAL RESEARCH

# Machine Learning-Based Discrimination of Cardiovascular Outcomes in Patients With Hypertrophic Cardiomyopathy



Tae-Min Rhee, MD,<sup>a,b,\*</sup> Yeon-Kyoung Ko, BA,<sup>c,d,\*</sup> Hyung-Kwan Kim, MD, PhD,<sup>a</sup> Seung-Bo Lee, PhD,<sup>d</sup> Bong-Seong Kim, PhD,<sup>e</sup> Hong-Mi Choi, MD, PhD,<sup>f</sup> In-Chang Hwang, MD, PhD,<sup>f</sup> Jun-Bean Park, MD, PhD,<sup>a</sup> Yeonyee E. Yoon, MD, PhD,<sup>f</sup> Yong-Jin Kim, MD, PhD,<sup>a</sup> Goo-Yeong Cho, MD, PhD<sup>f</sup>

## ABSTRACT

**BACKGROUND** Current risk stratification strategies for patients with hypertrophic cardiomyopathy (HCM) are limited to traditional methodologies.

**OBJECTIVES** The authors aimed to establish machine learning (ML)-based models to discriminate major cardiovascular events in patients with HCM.

**METHODS** We enrolled consecutive HCM patients from 2 tertiary referral centers and used 25 clinical and echocardiographic features to discriminate major adverse cardiovascular events (MACE), including all-cause death, admission for heart failure (HF-adm), and stroke. The best model was selected for each outcome using the area under the receiver operating characteristic curve (AUROC) with 20-fold cross-validation. After testing in the external validation cohort, the relative importance of features in discriminating each outcome was determined using the SHapley Additive exPlanations (SHAP) method.

**RESULTS** In total, 2,111 patients with HCM (age  $61.4 \pm 13.6$  years; 67.6% men) were analyzed. During the median 4.0 years of follow-up, MACE occurred in 341 patients (16.2%). Among the 4 ML models, the logistic regression model achieved the best AUROC of 0.800 (95% CI: 0.760–0.841) for MACE, 0.789 (95% CI: 0.736–0.841) for all-cause death, 0.798 (95% CI: 0.736–0.860) for HF-adm, and 0.807 (95% CI: 0.754–0.859) for stroke. The discriminant ability of the logistic regression model remained excellent when applied to the external validation cohort for MACE (AUROC = 0.768), all-cause death (AUROC = 0.750), and HF-adm (AUROC = 0.806). The SHAP analysis identified left atrial diameter and hypertension as important variables for all outcomes of interest.

**CONCLUSIONS** The proposed ML models incorporating various phenotypes from patients with HCM accurately discriminated adverse cardiovascular events and provided variables with high importance for each outcome. (JACC: Asia 2024;4:375–386) © 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 the <sup>a</sup>Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea; <sup>b</sup>Department of Internal Medicine, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, Republic of Korea; <sup>c</sup>Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul, Republic of Korea; <sup>d</sup>Department of Medical Informatics, Keimyung University School of Medicine, Daegu, Republic of Korea; <sup>e</sup>Department of Statistics and Actuarial Science, The Soongsil University, Seoul, Republic of Korea; and the <sup>f</sup>Cardiovascular Center and Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea. \*Drs Rhee and Ko 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 July 28, 2023; revised manuscript received November 7, 2023, accepted December 4, 2023.

## ABBREVIATIONS AND ACRONYMS

**AF** = atrial fibrillation

**AUROC** = area under the receiver operating characteristic curve

**HCM** = hypertrophic cardiomyopathy

**HF** = heart failure

**HF-adm** = heart failure admission

**LA** = left atrium

**LR** = logistic regression

**LV** = left ventricle

**LVESVI** = left ventricle end-systolic volume index

**ML** = machine learning

**SCD** = sudden cardiac death

**SHAP** = SHapley Additive exPlanations

**H**ypertrophic cardiomyopathy (HCM) is an inheritable myocardial disease with a prevalence of 1:500 in the general population.<sup>1</sup> HCM is a well-known leading cause of sudden cardiac death (SCD), especially in young individuals.<sup>2,3</sup> Thus, the current risk stratification approach for patients with HCM has primarily focused on SCD.<sup>4,5</sup> Given the advancing age of newly diagnosed HCM patients,<sup>6</sup> however, it is not surprising that their life expectancy and quality of life are largely determined by cardiovascular complications, such as heart failure (HF),<sup>7</sup> stroke,<sup>8</sup> or atrial fibrillation (AF).<sup>9</sup> Notwithstanding, well-validated prediction models for adverse cardiovascular events are hitherto rare in HCM.

Machine learning (ML)-based models containing multidimensional variables predict adverse events with more precision and generalizability than conventional risk predictors in various cardiovascular diseases.<sup>10</sup> By incorporating high-order and nonlinear interactions among variables, ML methods provide improved predictive ability compared with the standard regression techniques.<sup>11</sup> Therefore, several ML models for predicting unfavorable cardiovascular outcomes in patients with HCM have been suggested.<sup>12-14</sup> However, previous models have not been validated in external independent cohorts, substantially limiting the generalizability of the established models. Thus, we aimed to establish and validate data-driven ML-based models to discriminate major cardiovascular events using 2 independent largescale HCM cohorts. Further, we utilized an explainable ML method to provide novel insights from important features affecting the decision-making process of ML prediction models.

## METHODS

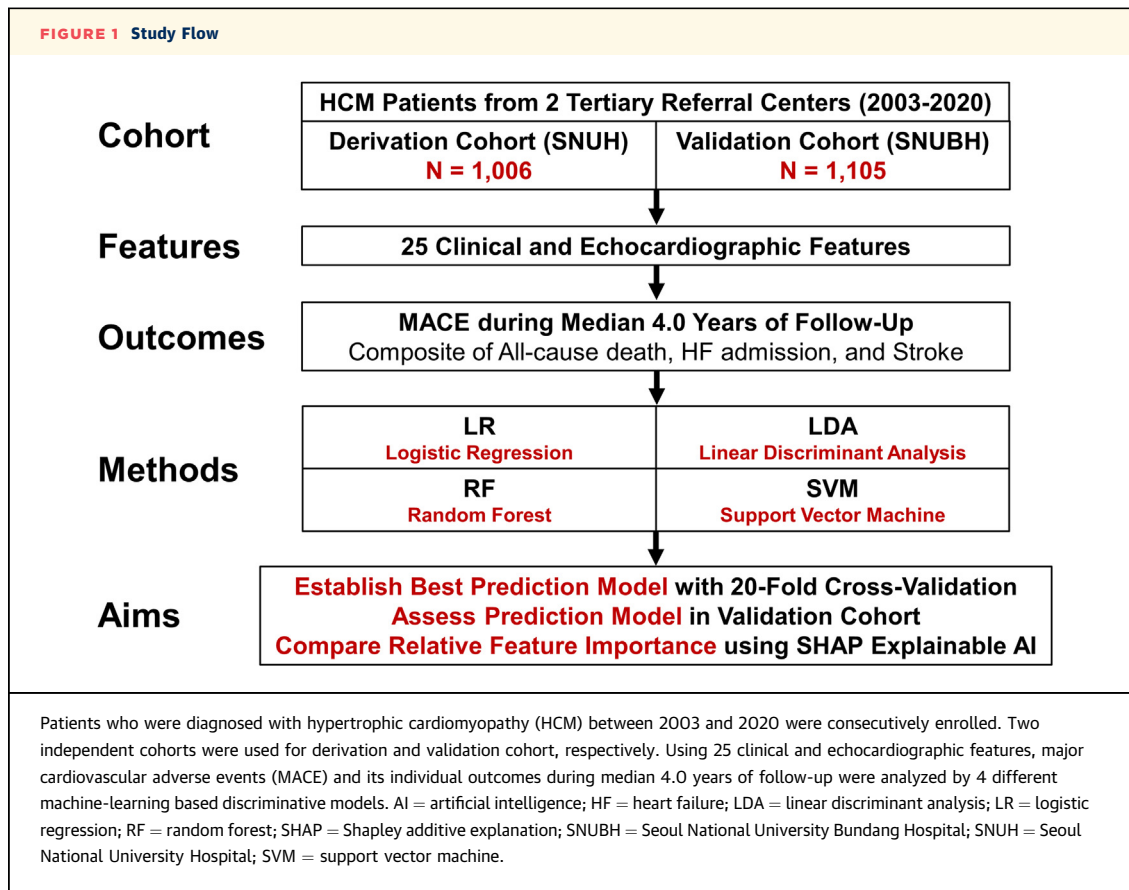
**STUDY DESIGN AND POPULATION.** The study population consisted of patients with HCM who received their first echocardiography examination between 2003 and 2020 at 2 tertiary referral hospitals in Korea (Seoul National University Hospital and Seoul National University Bundang Hospital) (Figure 1). HCM was defined as an increased left ventricular (LV) wall thickness (end-diastolic LV wall thickness  $\geq 15$  mm or  $\geq 13$  mm in individuals with a familial history of HCM), with LV hypertrophy unattributable to secondary causes such as hypertension or aortic stenosis.<sup>5</sup> In order to establish the ML models, the cohort from Seoul National University Hospital (n = 1,006) was assigned as the derivation cohort.

The Seoul National University Bundang Hospital cohort (n = 1,105) was the external validation cohort used to evaluate the predictive ability of the best prediction model. The study protocol was approved by the institutional review board of each hospital and conducted according to the principles of the Declaration of Helsinki. Written informed consent was waived due to the retrospective nature of the study.

**ECHOCARDIOGRAPHIC EXAMINATION.** Three distinct vendors (GE Medical Systems, Philips Healthcare, and Siemens Medical Solutions) were utilized for the echocardiographic examination. Dimensions of the LV and left atrium (LA) and LV end-diastolic wall thickness were measured on parasternal long-axis views. In addition, the LV ejection fraction was evaluated on the apical 4- and 2-chamber views using the biplane Simpson's method following the guidelines.<sup>15</sup> The LV outflow tract pressure gradient was measured at rest and with the Valsalva maneuver, with the maximum value acquired. Apical HCM was defined as pathological hypertrophy of the LV limited to the apical segments of the ventricle.

**FEATURE SELECTION AND DATA PREPROCESSING.** We first excluded 5 echocardiographic parameters (LV internal diameter at end-systole, LV end-diastolic volume index, stroke volume index, medial  $e'$ , and E velocity) from 30 features due to a multicollinearity issue. Because the proportion of missing values for each feature was  $<5\%$ , no feature was removed due to the excessive missing values. As a result, 15 clinical and 10 echocardiographic features were included in the analysis. The variables used in the final analysis are as follows: Age, sex, body mass index, hypertension, diabetes mellitus, dyslipidemia, history of HF, AF, history of stroke, coronary artery disease, valvular heart disease, history of cancer, family history of SCD, history of syncope, apical type HCM, use of beta-blocker, maximum LV wall thickness, maximum LV outflow tract pressure gradient, LV internal diameter at end-diastole, LV end-systolic volume index (LVESVi), LV ejection fraction, LA diameter, deceleration time,  $E/e'$ , and estimated pulmonary artery systolic pressure. The K nearest neighbor imputation method was used to fill in missing values to prevent the loss of important information and unstable model implementation. All results were analyzed based on the imputed dataset.

**STUDY OUTCOMES AND LONGITUDINAL FOLLOW-UP.** The outcomes of interest were MACE (defined as a composite of all-cause mortality, HF admission [HF-adm], and stroke) and the individual components. Using the National Death Registration Records of Korea (independently managed by the Korean



government), the vital status of all study participants was verified. HF-adm was defined as at least 1 episode of hospitalization owing to HF, which was clinically diagnosed based on aggravating symptoms and indications of congestion with volume overloads, including dyspnea and peripheral edema, or using diuretic agents for volume overload.<sup>16</sup> Stroke was a sudden neurologic impairment due to vascular pathologies of the brain, such as thromboembolism, hemorrhage, or ruptured aneurysm, that lasted over 24 hours. Dedicated research personnel gathered clinical outcomes by reviewing electronic health records and performing telephone interviews. From the initial echocardiographic examination date, participants were followed up until the event, the end of follow-up, or death, whichever occurred first. The median follow-up was 4.0 years (Q1-Q3: 1.6-7.6 years).

**DEVELOPMENT OF ML MODELS AND EXTERNAL VALIDATION OF THE BEST PREDICTION MODEL.**

In contrast to the traditional methodologies, ML-based methods utilize all provided variables through

multiple iteration processes to select the optimal model.<sup>17</sup> Consequently, ML-based methods can manage nonlinearity, high dimensionality, and variable interactions, delivering more accurate variable importance and superior predictive performance.<sup>18</sup> In this study, 4 ML classifiers—logistic regression (LR), linear discriminant analysis, random forest, and support vector machine—were used to construct the discriminative models for each target outcome. The 20-fold cross-validation method was used for the fine-tuning of the optimal model parameters in the derivation cohort. The performance of the established model in the derivation cohort was assessed using the area under the receiver operating characteristic curve (AUROC) and the calibration plot by deciles of predicted risk of event. Based on the AUROC values, the best performing model was chosen for each outcome. We then evaluated the discriminant ability of ML models in the external validation cohort. A sensitivity analysis was followed as a cross-over analysis with the roles of the derivation and validation cohorts switched. In addition, we performed a hold-out

**TABLE 1 Baseline Characteristics of Study Population**

	Derivation Cohort (n = 1,006)	Validation Cohort (n = 1,105)	P Value
<b>Demographic features</b>			
Age, y	61.6 ± 13.3	61.3 ± 13.8	0.630
Male	67.2 (676)	67.9 (750)	0.740
Body mass index, kg/m <sup>2</sup>	24.8 ± 3.3	24.9 ± 3.5	0.247
<b>Comorbidities</b>			
Hypertension	45.2 (455)	62.8 (694)	<0.001
Diabetes mellitus	19.1 (192)	22.7 (251)	0.041
Dyslipidemia	23.6 (237)	44.9 (496)	<0.001
History of heart failure	4.1 (41)	4.8 (53)	0.423
Atrial fibrillation	21.5 (216)	16.6 (183)	0.004
History of stroke	4.2 (42)	13.2 (146)	<0.001
Coronary artery disease	10.9 (110)	8.0 (88)	0.019
Valvular heart disease	3.9 (39)	4.9 (54)	0.259
History of cancer	11.0 (111)	13.0 (144)	0.159
<b>HCM-related risk factors</b>			
Familial history of SCD	7.6 (76)	7.1 (79)	0.721
History of syncope	13.3 (134)	7.6 (84)	<0.001
Apical type	43.2 (435)	38.6 (426)	0.029
Beta-blocker use	35.6 (358)	69.4 (767)	<0.001
<b>Echocardiographic parameters</b>			
Maximal LV wall thickness, mm	18.3 ± 4.0	18.0 ± 3.8	0.039
Maximal LVOT pressure gradient, mm Hg	13.9 ± 27.2	16.3 ± 31.3	0.068
LVEDVi, mL/m <sup>2</sup>	86.3 ± 19.1	81.2 ± 19.1	<0.001
LVESVi, mL/m <sup>2</sup>	30.7 ± 10.4	30.1 ± 11.9	0.230
LVEF, %	63.6 ± 6.9	63.8 ± 7.2	0.471
LA diameter, mm	45.8 ± 7.5	42.3 ± 7.6	<0.001
E velocity, m/s	0.6 ± 0.2	0.7 ± 0.2	<0.001
Deceleration time, msec	211.6 ± 90.5	205.3 ± 68.6	0.073
Medial e', cm/s	4.7 ± 1.6	5.5 ± 2.0	<0.001
E/e'	14.4 ± 6.9	13.5 ± 6.6	0.005
Estimated PASP, mm Hg	17.6 ± 17.7	29.5 ± 9.9	<0.001
Values are mean ± SD or % (n). HCM = hypertrophic cardiomyopathy; LA = left atrium; LV = left ventricle; LVEDVi = left ventricular end-diastolic volume index; LVEF = left ventricular ejection fraction; LVESVi = left ventricular end-systolic volume index; LVIDd = left ventricular internal dimension at end-diastole; LVIDs = left ventricular internal dimension at end-systole; LVOT = left ventricular outflow tract; PASP = pulmonary arterial systolic pressure; SCD = sudden cardiac death.			

cross-validation analysis by merging the cohorts into a single pooled dataset of 2,111 patients, which was then randomly split into training and test cohorts in a 7:3 manner.

**FEATURE IMPORTANCE ANALYSIS USING THE SHAPLEY ADDITIVE EXPLANATIONS METHOD.** We utilized the Shapley additive explanations (SHAP) method to rank the relative importance of each feature incorporated in the ML models with best discriminant ability established from the derivation cohort. Using the SHAP method, the Shapley values, indicating the contribution of each feature to the predictive probability of target outcomes, were estimated.<sup>19</sup> Considering the possible interactions among all features, the Shapley value of a feature represents the effect of deleting the feature from the prediction

model.<sup>20</sup> Model development was done using Python software version 3.7.6 (Python Software Foundation) and Scikit Learn software version 1.0.2 (scikit-learn Project). The SHAP analysis results were reported as radar plots and the SHAP summary plots, depicting scaled importance (the relative importance of a feature scaled with respect to the feature with the greatest relative importance value) of each feature.

**STATISTICAL ANALYSIS.** The chi-square test was utilized to compare categorical variables expressed as numbers and relative frequencies (percentages). Continuous variables were reported as mean ± SD. We utilized the Kolmogorov-Smirnov test to determine whether the continuous variables had a normal distribution. Student's *t*-test was used to analyze differences between continuous characteristics with a normal distribution; otherwise, the Mann-Whitney *U* test was employed. All probability values were 2-sided, and *P* values <0.05 were considered significant. Analyses were conducted using Stata software version 17.0 (StataCorp).

## RESULTS

**CHARACTERISTICS AND CLINICAL OUTCOMES OF THE STUDY POPULATION.** The baseline characteristics of the derivation and validation cohorts are presented in **Table 1**. In total, 2,111 patients (derivation cohort n = 1,006; validation cohort n = 1,105) with HCM (mean age 61.4 ± 13.6 years; 67.6% men) were analyzed. The distribution of age, sex, and body mass index were comparable between the 2 cohorts. Among HCM-related factors, the derivation cohort had a higher proportion of patients with a history of syncope and apical type HCM. The proportion of patients on beta-blockers was higher in the validation cohort. The incidence rate of each outcome is presented in **Table 2**. During the median follow-up of 4.0 years, 3.72 per 100 person-years in the derivation cohort and 3.35 per 100 person-years in the validation cohort had MACE events.

**DEVELOPMENT AND EXTERNAL VALIDATION OF THE ML PREDICTION MODEL FOR CLINICAL OUTCOMES.** The ability of 4 ML models to discriminate clinical outcomes is depicted in **Figure 2**. The 20-fold cross-validated AUROC was 0.70 to 0.80 for each ML model, indicating excellent discriminant ability for all target outcomes (**Table 3**). The LR model demonstrated the best performance for all outcomes of interest among the 4 ML models. The 20-fold cross-validated AUROC was 0.800 (95% CI: 0.760-0.841) for MACE, 0.789 (95% CI: 0.736-0.841) for all-cause death, 0.798 (95% CI: 0.736-0.860) for HF-adm, and 0.807 (95% CI: 0.754-0.859) for stroke. Calibration bar

plots by deciles of the observed vs predicted risk estimated using the LR model are presented for each outcome in **Figure 3**. Overall, the LR model-predicted probability corresponded with the observed probability of each event in the derivation cohort. **Figure 4** and **Table 4** present the AUROC curves and detailed metrics of the LR model performances in the external validation cohort. When applied to the validation cohort, the performance of the LR model remained excellent for MACE (AUROC = 0.771), all-cause death (AUROC = 0.759), and HF-adm (AUROC = 0.807), except for stroke (AUROC = 0.694).

**SENSITIVITY ANALYSIS.** Sensitivity analyses were performed to replicate the results. As a cross-over analysis, LR-based ML models were developed from the Seoul National University Bundang Hospital cohort (validation cohort in the main analysis) and subsequently tested in the Seoul National University Hospital cohort (derivation cohort in the main analysis) (**Supplemental Table 1**). The discriminant ability for MACE, all-cause death, and HF-adm was 0.759 to 0.790 when applied to the cross-over validation cohort. Similar to the main result, the LR model performance decreased to AUROC of 0.655 for stroke prediction when externally validated.

A hold-out cross-validation analysis was performed as another sensitivity analysis (**Supplemental Table 2**). Random splitting of the merged total dataset in a 7:3 manner yielded training sets (n = 1,478) and test sets (n = 633) for analysis. The LR-based ML model established in the training set yielded an AUROC of 0.757 to 0.828. The model performance remained excellent in the test set, with an AUROC of 0.778 to 0.824 for all outcomes.

To assess the impact of the missing data, a complete data analysis without imputation was performed (**Supplemental Tables 3 and 4, Supplemental Figure 1**). Both the main result and the SHAP analysis results were consistent, demonstrating the robustness of established ML models across all outcomes.

**SHAP FEATURE IMPORTANCE ANALYSIS.** The relative importance of the top 8 features as significant factors for each clinical outcome was determined using Shapley values (**Figure 5**). The high-rank variables varied depending on the outcomes of interest. Variables with the highest importance were age (for MACE and all-cause death), AF (for HF-adm), and apical HCM (as a protective predictor for stroke). Increased LA diameter and hypertension, which ranked highly in the ML models, strongly associated with all 4 outcomes. Increased LVESVi and cancer history were substantially linked with more than 2 individual outcomes (for LVESVi, all-cause death,

**TABLE 2 Incidence Rate of Study Outcomes**

	Derivation Cohort (n = 1,006)	Validation Cohort (n = 1,105)
MACE		
Incidence rate, per 100 person-years	3.72	3.35
All-cause death		
Incidence rate, per 100 person-years	2.46	1.11
HF admission		
Incidence rate, per 100 person-years	1.51	1.24
Stroke		
Incidence rate, per 100 person-years	1.43	1.42

HF = heart failure; MACE = major adverse cardiovascular event(s).

and HF-adm; for cancer history, all-cause death, and stroke) and MACE.

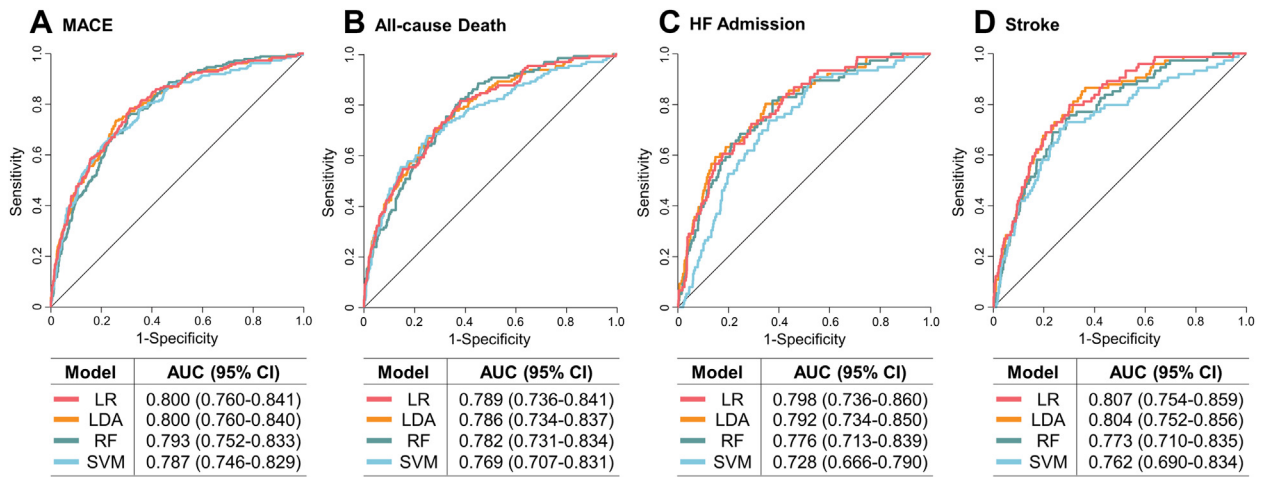
The variables selected through the traditional stepwise logistic regression method were compared to the variables identified using the SHAP method (**Supplemental Table 5**). One consistent finding was the significance of the LA diameter across all outcomes. However, there were notable differences in the lists of selected variables for each outcome.

**DISCUSSION**

In this study, we developed ML-based discriminative models using independent consecutive HCM cohorts from 2 tertiary referral centers, focusing on the risk of MACE, including all-cause death, HF-adm, and stroke (**Central Illustration**). Our findings can be summarized as follows. First, among the 4 ML models, the LR-based ML algorithm had the best discriminant ability for all 4 outcomes. Second, LR models for all outcomes were well-calibrated and maintained good discriminant ability when applied to the external validation cohort, except for stroke. Third, the relative importance of clinical and echocardiographic parameters in discriminating each outcome was determined using the SHAP analysis. We observed that LA diameter and hypertension had substantial importance in discriminating all four outcomes.

In patients with HCM, major cardiovascular complications, such as HF or stroke, significantly affect their quality of life and prognosis.<sup>21</sup> Patients with HCM are at a substantially greater risk of mortality owing to cardiovascular diseases than those without.<sup>4</sup> In contrast to SCD, studies on prediction models and predictors for adverse cardiovascular outcomes in HCM are anecdotal. Current prediction models depend on single-center observational data, established based on traditional methodologies with hand-crafted criteria.<sup>2</sup> These models may be vulnerable and perform poorly when applied to the new data.<sup>2</sup> By

**FIGURE 2 Performance of 4 ML Models for Each Clinical Outcome**



Predictive ability of 4 machine learning (ML) discriminative models for (A) MACE, (B) death, (C) HF admission, and (D) stroke are presented as receiver-operating characteristic (ROC) curves and compared by the area under the receiver-operating curve (AUROC) with 95% CI. Abbreviations as in Figure 1.

thoroughly incorporating multidimensional data and factors that interact in linear and nonlinear manners, the ML-based methodology may provide a model with significantly enhanced prediction performance.<sup>22</sup> Indeed, ML models to predict ventricular arrhythmia,<sup>14</sup> HF,<sup>13,14</sup> or composite cardiovascular

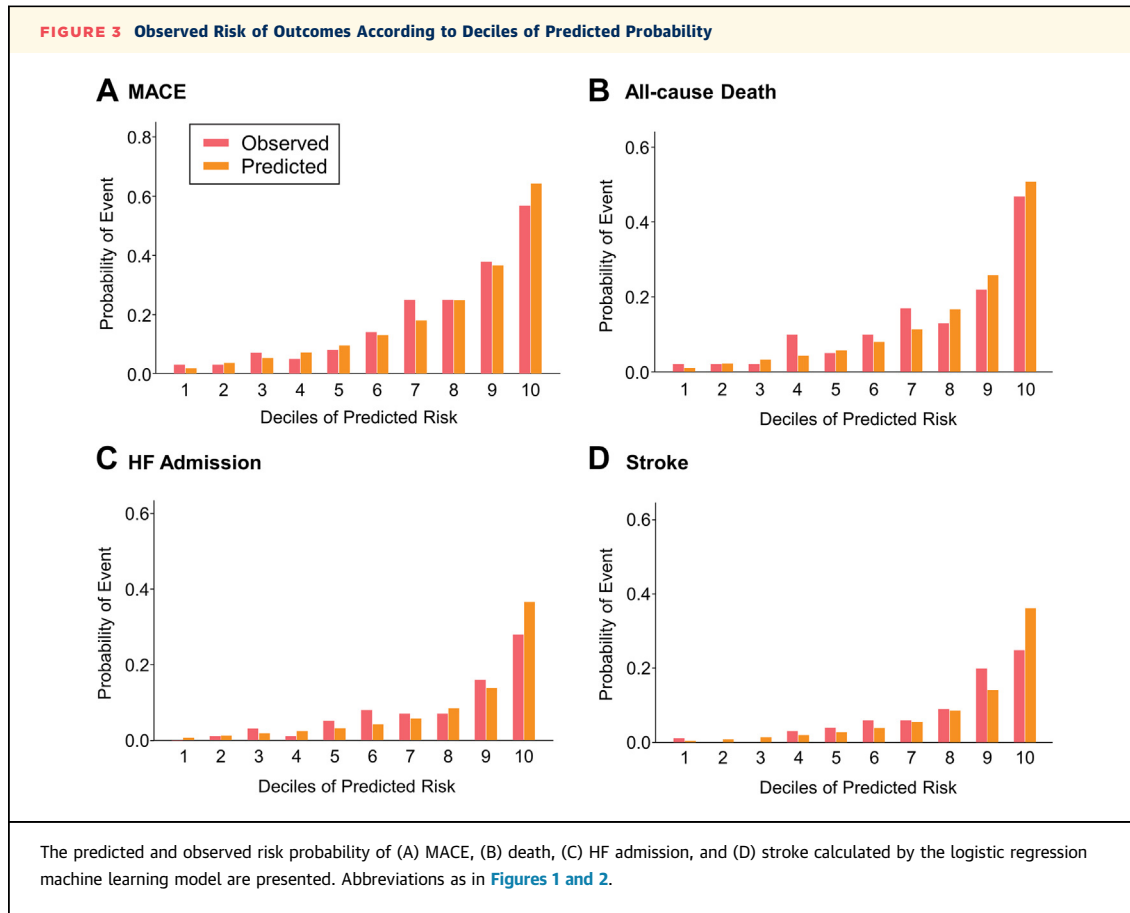
events<sup>12</sup> were recently presented in HCM populations. However, all previous studies only offered an ML model based on single-center data that had not been verified in an external cohort. This is a critical drawback that substantially limits the applicability of the established models. There is still room for

**TABLE 3 Predictive Performance of Machine Learning-Based Models by 20-Fold Cross-Validation for Each Clinical Outcome in Derivation Cohort**

	AUC	Sensitivity	Specificity	PPV	NPV
<b>MACE</b>					
Logistic regression	0.800 (0.760-0.841)	0.655 (0.592-0.718)	0.795 (0.740-0.851)	0.460 (0.393-0.527)	0.912 (0.899-0.926)
Linear discriminant analysis	0.800 (0.760-0.840)	0.659 (0.595-0.722)	0.794 (0.738-0.850)	0.456 (0.398-0.515)	0.914 (0.901-0.926)
Random forest	0.793 (0.752-0.833)	0.771 (0.705-0.837)	0.688 (0.608-0.768)	0.394 (0.334-0.453)	0.934 (0.921-0.946)
Support vector machine	0.787 (0.746-0.829)	0.666 (0.597-0.735)	0.788 (0.718-0.858)	0.488 (0.396-0.580)	0.914 (0.900-0.927)
<b>All-cause death</b>					
Logistic regression	0.789 (0.736-0.841)	0.698 (0.599-0.796)	0.720 (0.645-0.795)	0.344 (0.249-0.438)	0.943 (0.929-0.958)
Linear discriminant analysis	0.786 (0.734-0.837)	0.667 (0.561-0.772)	0.747 (0.668-0.827)	0.365 (0.270-0.459)	0.941 (0.927-0.955)
Random forest	0.782 (0.731-0.834)	0.707 (0.627-0.787)	0.704 (0.635-0.772)	0.288 (0.236-0.340)	0.943 (0.931-0.956)
Support vector machine	0.769 (0.707-0.831)	0.608 (0.547-0.734)	0.785 (0.730-0.818)	0.308 (0.272-0.345)	0.934 (0.923-0.951)
<b>HF admission</b>					
Logistic regression	0.798 (0.736-0.860)	0.671 (0.621-0.721)	0.675 (0.575-0.775)	0.212 (0.118-0.305)	0.954 (0.936-0.972)
Linear discriminant analysis	0.792 (0.734-0.850)	0.633 (0.557-0.710)	0.739 (0.652-0.827)	0.257 (0.157-0.357)	0.960 (0.951-0.969)
Random forest	0.776 (0.713-0.839)	0.550 (0.448-0.652)	0.789 (0.716-0.862)	0.281 (0.167-0.395)	0.956 (0.948-0.964)
Support vector machine	0.728 (0.666-0.790)	0.596 (0.491-0.701)	0.676 (0.584-0.768)	0.171 (0.116-0.226)	0.954 (0.946-0.962)
<b>Stroke</b>					
Logistic regression	0.807 (0.754-0.859)	0.600 (0.509-0.691)	0.758 (0.676-0.840)	0.287 (0.156-0.418)	0.960 (0.952-0.967)
Linear discriminant analysis	0.804 (0.752-0.856)	0.596 (0.508-0.684)	0.769 (0.697-0.841)	0.274 (0.160-0.389)	0.960 (0.952-0.968)
Random forest	0.773 (0.710-0.835)	0.671 (0.615-0.727)	0.693 (0.617-0.768)	0.191 (0.121-0.262)	0.962 (0.955-0.969)
Support vector machine	0.762 (0.690-0.834)	0.588 (0.498-0.677)	0.791 (0.737-0.845)	0.243 (0.149-0.337)	0.961 (0.953-0.968)

Values in ( ) are 95% CI.

AUC = area under the receiver operating characteristic curve; NPV = negative predictive value; PPV = positive predictive value; other abbreviations as in Table 2.



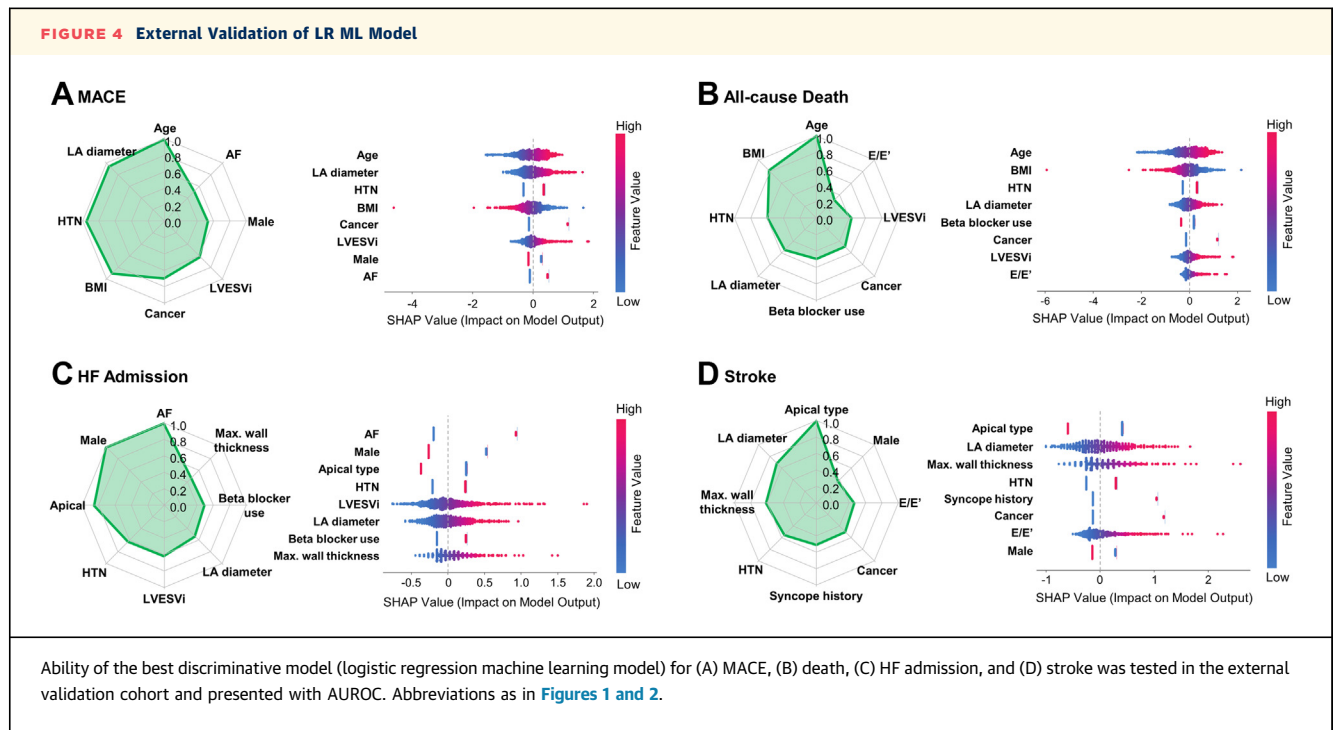
improvement regarding the precision and transferability of these ML models, along with a need for tailored approaches in HCM patient care.

Our study has the following methodological strengths: 1) for the first time, we established the generalizability of a model developed from 2 independent HCM cohorts (each with over 1,000 patients); 2) we analyzed and compared 4 ML approaches frequently used in classification, including LR, linear discriminant analysis, random forest, and support vector machine, to determine the most appropriate ML prediction model for our data distribution; and 3) model training with 20-fold cross-validation. Moreover, sensitivity analyses using cross-over or hold-out cross-validation further improved and consolidated the reliability of our results.

The LR ML model provided the best discriminant ability for all 4 outcomes of interest in this study population. However, the lowest AUROCs of ML models in our study were still comparable to those of prior research, falling in the range of 0.73 to

0.79. Consequently, the data structure and distribution from the derivation and validation cohorts used in this study were generally suitable for developing and applying ML-based models. This indicates that our data may be utilized to train and validate other ML-based models with novel methodologies.

Notably, the discriminant performance of the model for stroke dropped significantly compared with other outcomes during external validation. A discrepancy in the baseline distribution between the derivation and validation cohorts might have contributed to this result. In particular, features directly associated with the occurrence of stroke, such as previous history of stroke, prevalence of AF, and LA diameter in echocardiography examination, significantly differed between 2 cohorts. Furthermore, because cohorts used in this study mainly focused on HCM and its cardiovascular outcomes, unmeasured factors that may significantly affect stroke occurrence (such as carotid artery stenosis,



smoking status, alcohol intake, or physical activity)<sup>23</sup> may not have been considered in our models. To further improve the efficacy of the stroke discriminative models, subsequent studies using cohorts with various phenotypes are helpful.

The ML model is a black-box system, making it challenging to comprehend decision-making. This results in limited implications on clinical practice despite the enhanced discriminant ability.<sup>24</sup> To overcome these limitations, we used the SHAP method—an explainable AI methodology based on game theory that indexes and communicates the degree of the relative contribution of features to the discriminative model via the Shapley value. The SHAP analysis demonstrated that the relative importance of variables changes meaningfully based on the outcomes. These outcome-specific important

features should be considered in further studies with larger, more comprehensive predictive models for cardiovascular outcomes in HCM patients. The significant discrepancies between the traditional stepwise variable selection and ML-based SHAP approach highlight an opportunity to update significant predictors considering high-dimensional intervariable interactions, and may provide novel clinical insights as well as new treatment targets.

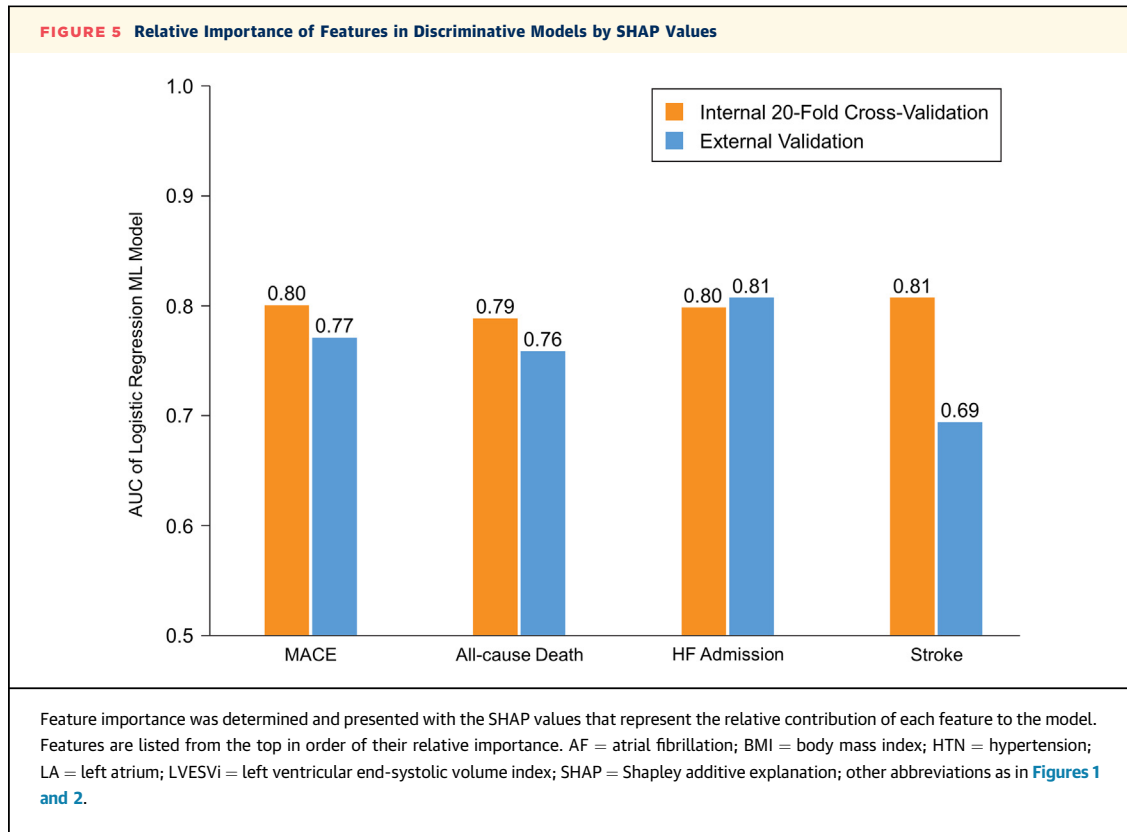
The SHAP analysis revealed that LA diameter and hypertension were common significant factors of the 4 outcomes of interest. Although routinely measured in echocardiography examinations, the size of the LA cavity is often overlooked in clinical practice. As a representative, chronic marker for the diastolic function of the LV, the LA size should be considered

**TABLE 4 Predictive Performance of Machine Learning-Based Logistic Regression Model for Each Clinical Outcome in Validation Cohort**

	AUC	Sensitivity	Specificity	PPV	NPV
MACE	0.771	0.705	0.737	0.306	0.938
All-cause death	0.759	0.746	0.657	0.102	0.980
HF admission	0.807	0.700	0.818	0.181	0.979
Stroke	0.694	0.912	0.380	0.088	0.985

Abbreviations as in [Tables 2 and 3](#).





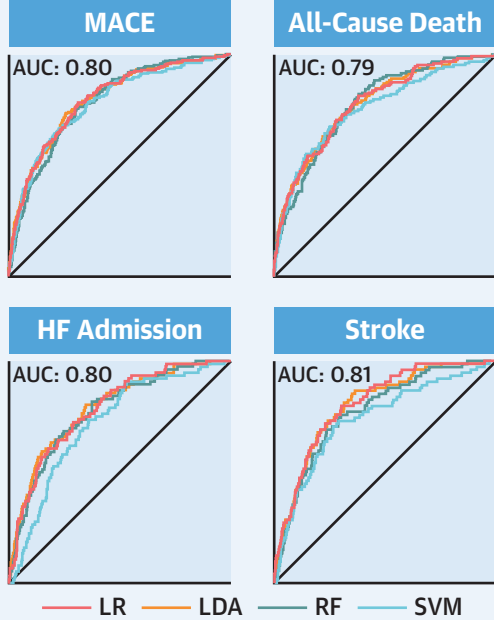
an essential imaging index for long-term prognosis in patients with HCM.<sup>25</sup> It is crucial to investigate how the LA reservoir strain—recently reported as a novel parameter with a strong predictive ability for incident HF in patients with HCM—may influence the ML-based model enhancement.<sup>26</sup> Hypertension is the leading cause of cardiovascular disease, substantially impacting mortality, HF, and stroke.<sup>27</sup> In our study cohorts, 54.4% of patients diagnosed with HCM had concomitant hypertension. This proportion is higher than that in studies recruiting a relatively young Western HCM population.<sup>12,13</sup> However, it is comparable to Japanese data<sup>28</sup> or the Korean national health insurance database,<sup>1</sup> reflecting the different characteristics of the study population. Uncontrolled hypertension enhances LV hypertrophy and hastens unfavorable remodeling of the myocardium.<sup>29</sup> Considering the continually aging HCM population,<sup>6</sup> it is necessary to carefully examine hypertension in patients with HCM at diagnosis and during follow-up, and to make every effort to manage blood pressure adequately. Furthermore, the LV chamber size and the patient’s malignancy history, which were noted

as additional key determinants in the occurrence of unfavorable cardiovascular events, including mortality, should not be ignored at the time of HCM diagnosis, especially in the contemporary era of the enhanced HCM care.

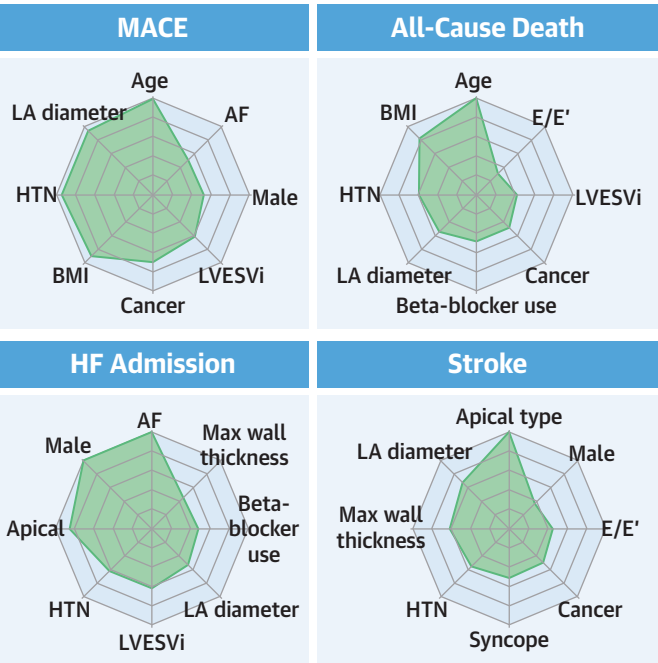
Notably, the apical type HCM was the most important factor inversely correlated with the incident stroke. Similarly, a Japanese HCM cohort study observed that apical HCM is linked with a lower incidence of thromboembolic events, including stroke.<sup>30</sup> Apical HCM is characterized by less severe diastolic dysfunction and a smaller extent of myocardial fibrosis than HCM with septal hypertrophy.<sup>31</sup> In this study, patients with non-apical HCM had a higher E/e’ ratio ( $14.7 \pm 7.4$  vs  $12.9 \pm 5.5$ ) despite similar sizes of LA cavity than those with apical type, suggesting a higher chance of LV diastolic dysfunction in patients with non-apical HCM. Furthermore, patients with HCM with a greater extent of septal hypertrophy were prone to AF, possibly explaining the higher risk of stroke in patients with non-apical HCM.<sup>32</sup> Because our data are limited to provide a clear explanation for causal relationship, this finding

**CENTRAL ILLUSTRATION** Machine Learning-Based Discrimination of Cardiovascular Outcomes in Hypertrophic Cardiomyopathy Patients

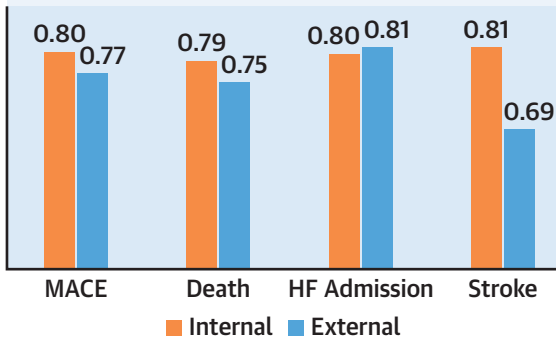
**20-Fold Cross-Validated Machine-Learning Models**



**SHAP-Value Based Feature Importance Ranking**



**External Validation of Best Discriminative Model**



Precise Risk Discrimination for Major Outcomes in Hypertrophic Cardiomyopathy Patients

Guidance for Clinical Practice Focusing on More Important Clinical/Imaging Parameters

Rhee T-M, et al. JACC: Asia. 2024;4(5):375-386.

This study aimed to develop and validate machine learning (ML)-based models for discriminating major cardiovascular events in hypertrophic cardiomyopathy (HCM) patients, using data from 2 largescale HCM cohorts from independent tertiary referral centers. Among the 4 ML models, the logistic regression-based ML algorithm had the best discriminant ability for all 4 outcomes. The relative importance of clinical and echocardiographic parameters in discriminating each outcome was determined using the SHapley Additive exPlanations (SHAP) analysis. We observed that left atrial (LA) diameter and hypertension (HTN) had substantial importance in predicting all 4 outcomes in patients with HCM. AF = atrial fibrillation; AUC = area under the curve; BMI = body mass index; HF = heart failure; LDA = linear discriminant analysis; LR = logistic regression; LVESi = left ventricular end-systolic volume index; MACE = major cardiovascular adverse event(s); RF = random forest; SVM = support vector machine.

is currently hypothesis-generating at best. Future prospective studies should validate the risk of stroke according to the type of HCM and reveal its mechanistic background.

**STUDY LIMITATIONS.** First, this study was based on real-world observational Asian cohort data. Therefore, the generalizability of the current ML models should be reinvestigated internationally using external HCM cohorts. Our established ML model pipeline can also be applied for validation within specific populations by age, sex, or underrepresented groups in future research. Second, we used ML classifiers that did not account for the time variable and, thus, did not consider the longitudinal aspect of the cohort study. However, our primary objective in this research was to compare various well-established and widely used ML classifiers. Third, the short follow-up periods and relatively few events for each individual outcome may limit the statistical power of this study. Fourth, the black box nature of the ML models limits the interpretability, which could be mitigated by utilizing the SHAP method in this study to elucidate the influence and importance of individual variables. Fifth, this study did not include the detailed assessment of HCM disease severity or risk stratification for HCM. Sixth, our data did not include parameters from advanced imaging techniques or functional tests, such as LV global longitudinal strain, LA reservoir strain, cardiac magnetic resonance imaging, treadmill test, or Holter monitoring. However, advanced imaging technique (eg, cardiac magnetic resonance) has its own limitations such as long acquisition time, difficult accessibility, and challenges in real-time evaluation. It is important that we incorporated the most commonly used clinical and echocardiographic parameters to pursue high adaptability of the established model.

## CONCLUSIONS

Applying ML approaches to models incorporating a wide variety of phenotypes improved the ability to discriminate major cardiovascular events and

provided features of high importance for each outcome in HCM patients. If clinically applicable, it can be helpful to early recognize and manage high-risk subsets of HCM patients.

## FUNDING SUPPORT AND AUTHOR DISCLOSURES

This study was supported by a research grant from the Seoul National University Research fund (no. 800-20210548). The funding source did not have any involvement on the study design, collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication. The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

**ADDRESS FOR CORRESPONDENCE:** Dr Hyung-Kwan Kim, Division of Cardiology, Department of Internal Medicine, Seoul National University College of Medicine, Cardiovascular Center, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea. E-mail: [cardiman73@gmail.com](mailto:cardiman73@gmail.com) or [hkim73@snu.ac.kr](mailto:hkim73@snu.ac.kr). OR Dr Seung-Bo Lee, Department of Medical Informatics, Keimyung University School of Medicine, 1095 Dalgubeoldae-ro, Dalseo-Gu, Daegu 42601, Republic of Korea. E-mail: [koreateam23@gmail.com](mailto:koreateam23@gmail.com).

## PERSPECTIVES

### COMPETENCY IN PATIENT CARE AND PROCEDURAL

**SKILLS:** We developed ML models for discriminating all-cause death, admission for heart failure, and stroke in HCM patients using data from 2 referral centers. ML-based discriminative models enhanced the accuracy of predicting major cardiovascular events in HCM patients, identifying important features for each outcome. Early recognition and management of high-risk HCM subsets could be facilitated by applying these models.

**TRANSLATIONAL OUTLOOK:** The findings of this study, which were derived from Asian data, underscore the importance of validating the results in cohorts representing diverse ethnic backgrounds, while also emphasize the necessity for refining the discriminative model by incorporating data from advanced cardiovascular imaging or functional test.

## REFERENCES

1. Moon I, Lee SY, Kim HK, et al. Trends of the prevalence and incidence of hypertrophic cardiomyopathy in Korea: a nationwide population-based cohort study. *PLoS One*. 2020;15:e0227012.
2. O'Mahony C, Jichi F, Pavlou M, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J*. 2014;35:2010-2020.
3. Lee HJ, Kim J, Chang SA, Kim YJ, Kim HK, Lee SC. Major clinical issues in hypertrophic cardiomyopathy. *Korean Circ J*. 2022;52:563-575.
4. Kwon S, Kim HK, Kim B, et al. Comparison of mortality and cause of death between adults with and without hypertrophic cardiomyopathy. *Sci Rep*. 2022;12:6386.
5. Lee HJ, Kim HK, Lee SC, et al. Supplementary role of left ventricular global longitudinal strain for predicting sudden cardiac death in hypertrophic cardiomyopathy. *Eur Heart J Cardiovasc Imaging*. 2022;23:1108-1116.
6. Canepa M, Fumagalli C, Tini G, et al. Temporal trend of age at diagnosis in hypertrophic cardiomyopathy: an analysis of the International

- Sarcomeric Human Cardiomyopathy Registry. *Circ Heart Fail*. 2020;13:e007230.
7. Choi YJ, Kim HK, Hwang IC, et al. Prognosis of patients with hypertrophic cardiomyopathy and low-normal left ventricular ejection fraction. *Heart*. 2023;109(10):771-778. <https://doi.org/10.1136/heartjnl-2022-321853>
  8. Choi YJ, Kim B, Rhee TM, et al. Augmented risk of ischemic stroke in hypertrophic cardiomyopathy patients without documented atrial fibrillation. *Sci Rep*. 2022;12:15785.
  9. Lee HJ, Kim HK, Kim M, et al. Clinical impact of atrial fibrillation in a nationwide cohort of hypertrophic cardiomyopathy patients. *Ann Transl Med*. 2020;8:1386.
  10. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular event prediction by machine learning: the Multi-Ethnic Study of Atherosclerosis. *Circ Res*. 2017;121:1092-1101.
  11. Shah SJ, Katz DH, Selvaraj S, et al. Phenotyping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015;131:269-279.
  12. Kochav SM, Raita Y, Fifer MA, et al. Predicting the development of adverse cardiac events in patients with hypertrophic cardiomyopathy using machine learning. *Int J Cardiol*. 2021;327:117-124.
  13. Fahmy AS, Rowin EJ, Manning WJ, Maron MS, Nezafat R. Machine learning for predicting heart failure progression in hypertrophic cardiomyopathy. *Front Cardiovasc Med*. 2021;8:647857.
  14. Smole T, Zunkovic B, Piculin M, et al. A machine learning-based risk stratification model for ventricular tachycardia and heart failure in hypertrophic cardiomyopathy. *Comput Biol Med*. 2021;135:104648.
  15. 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. *J Am Soc Echocardiogr*. 2015;28:1-39.e14.
  16. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). *Eur Heart J*. 2016;37:2129-2200.
  17. Pfob A, Lu SC, Sidey-Gibbons C. Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. *BMC Med Res Methodol*. 2022;22:282.
  18. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. 2019;19:64.
  19. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56-67.
  20. Wang K, Tian J, Zheng C, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med*. 2021;137:104813.
  21. Maron BJ, Rowin EJ, Casey SA, et al. Risk stratification and outcome of patients with hypertrophic cardiomyopathy  $\geq 60$  years of age. *Circulation*. 2013;127:585-593.
  22. Krittanawong C, Johnson KW, Rosenson RS, et al. Deep learning for cardiovascular medicine: a practical primer. *Eur Heart J*. 2019;40:2058-2073.
  23. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol*. 2021;20(10):795-820. [https://doi.org/10.1016/S1474-4422\(21\)00252-0](https://doi.org/10.1016/S1474-4422(21)00252-0)
  24. Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. *Can J Cardiol*. 2022;38:204-213.
  25. Nagueh SF, Smiseth OA, Appleton CP, et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr*. 2016;29:277-314.
  26. Lee HJ, Kim HK, Rhee TM, et al. Left atrial reservoir strain-based left ventricular diastolic function grading and incident heart failure in hypertrophic cardiomyopathy. *Circ Cardiovasc Imaging*. 2022;15:e013556.
  27. Unger T, Borghi C, Charchar F, et al. 2020 International Society of Hypertension global hypertension practice guidelines. *Hypertension*. 2020;75:1334-1357.
  28. Higuchi S, Minami Y, Shoda M, et al. Effect of renal dysfunction on risk of sudden cardiac death in patients with hypertrophic cardiomyopathy. *Am J Cardiol*. 2021;144:131-136.
  29. Gonzalez A, Ravassa S, Lopez B, et al. Myocardial remodeling in hypertension. *Hypertension*. 2018;72:549-558.
  30. Haruki S, Minami Y, Hagiwara N. Stroke and embolic events in hypertrophic cardiomyopathy: risk stratification in patients without atrial fibrillation. *Stroke*. 2016;47:936-942.
  31. Kim EK, Lee SC, Hwang JW, et al. Differences in apical and non-apical types of hypertrophic cardiomyopathy: a prospective analysis of clinical, echocardiographic, and cardiac magnetic resonance findings and outcome from 350 patients. *Eur Heart J Cardiovasc Imaging*. 2016;17:678-686.
  32. Park KM, Im SI, Kim EK, et al. Atrial fibrillation in hypertrophic cardiomyopathy: is the extent of septal hypertrophy important? *PLoS One*. 2016;11:e0156410.

---

**KEY WORDS** heart failure, hypertrophic cardiomyopathy, machine learning, prediction model, prognosis

---

**APPENDIX** For a supplemental figure and tables, please see the online version of this paper.

---

