JACC: CLINICAL ELECTROPHYSIOLOGY VOL. 9, NO. 8, 2023

PUBLISHED BY ELSEVIER

ORIGINAL RESEARCH

ARRHYTHMIA PROGNOSIS - AI/ML

A Novel ECG-Based Deep Learning Algorithm to Predict Cardiomyopathy in Patients With Premature Ventricular Complexes

Joshua Lampert, MD,[a,*] Akhil Vaid, MD,[b,*] William Whang, MD,[a] Jacob Koruth, MBBS,[a] Marc A. Miller, MD,[a] Marie-Noelle Langan, MD,[a] Daniel Musikantow, MD,[a] Mohit Turagam, MD,[a] Abhishek Maan, MD,[a] Iwanari Kawamura, MD,[a] Srinivas Dukkipati, MD,[a] Girish N. Nadkarni, MD,[b,c,†] Vivek Y. Reddy, MD[a,†]

ABSTRACT

BACKGROUND Premature ventricular complexes (PVCs) are prevalent and, although often benign, they may lead to PVC-induced cardiomyopathy. We created a deep-learning algorithm to predict left ventricular ejection fraction (LVEF) reduction in patients with PVCs from a 12-lead electrocardiogram (ECG).

OBJECTIVES This study aims to assess a deep-learning model to predict cardiomyopathy among patients with PVCs.

METHODS We used electronic medical records from 5 hospitals and identified ECGs from adults with documented PVCs. Internal training and testing were performed at one hospital. External validation was performed with the others. The primary outcome was first diagnosis of LVEF ≤40% within 6 months. The dataset included 383,514 ECGs, of which 14,241 remained for analysis. We analyzed area under the receiver operating curves and explainability plots for representative patients, algorithm prediction, PVC burden, and demographics in a multivariable Cox model to assess independent predictors for cardiomyopathy.

RESULTS Among the 14,241-patient cohort (age 67.6 ± 14.8 years; female 43.8%; White 29.5%, Black 8.6%, Hispanic 6.5%, Asian 2.2%), 22.9% experienced reductions in LVEF to ≤40% within 6 months. The model predicted reductions in LVEF to ≤40% with area under the receiver operating curve of 0.79 (95% CI: 0.77-0.81). The gradient weighted class activation map explainability framework highlighted the sinus rhythm QRS complex-ST segment. In patients who underwent successful PVC ablation there was a post-ablation improvement in LVEF with resolution of cardiomyopathy in most (89%) patients.

CONCLUSIONS Deep-learning on the 12-lead ECG alone can accurately predict new-onset cardiomyopathy in patients with PVCs independent of PVC burden. Model prediction performed well across sex and race, relying on the QRS complex/ST-segment in sinus rhythm, not PVC morphology. (J Am Coll Cardiol EP 2023;9:1437–1451) © 2023 by the American College of Cardiology Foundation.

From the [a]Helmsley Electrophysiology Center, Icahn School of Medicine at Mount Sinai, New York, New York, USA; [b]The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA; and [c]The Division of Data Driven and Digital Medicine (D3M), The Charles Bronfman Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. *Drs Lampert and Vaid contributed equally to this work and are joint first authors. †Drs Nadkarni and Reddy contributed equally to this work and are joint 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 12, 2023; revised manuscript received May 17, 2023, accepted May 24, 2023.

ISSN 2405-500X/$36.00 https://doi.org/10.1016/j.jacep.2023.05.025

ABBREVIATIONS AND ACRONYMS

AUROC = area under the receiver operating characteristic curve

AUPRC = area under the precision recall curve

GradCAM = gradient weighted class activation map

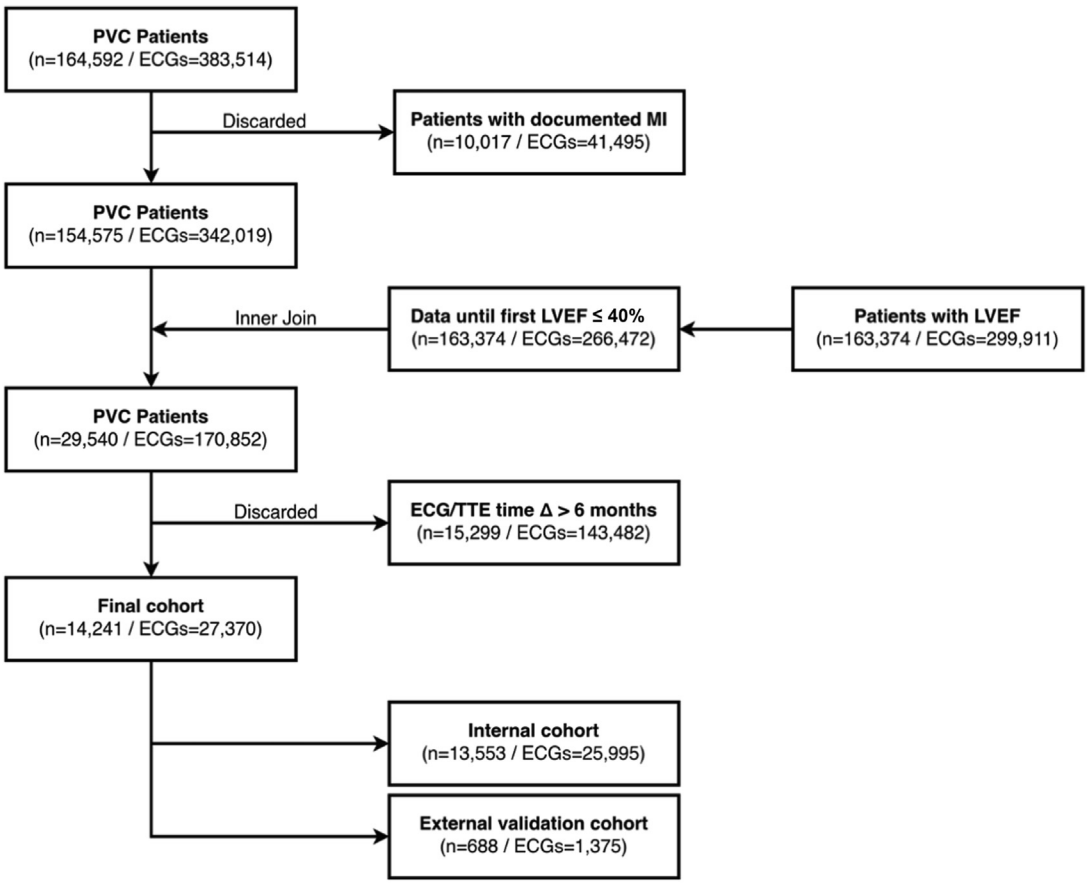
LVEF = left ventricular ejection fraction

PVC = premature ventricular contraction

Premature ventricular complexes (PVCs) are early depolarizations of the ventricular myocardium and are prevalent in 1% to 4% of the general adult population on 12-lead electrocardiography (ECG) and 40% to 75% on Holter monitoring.[1,2] Although initially considered benign, the 1990s saw the advent of our understanding of the concept of PVC-induced cardiomyopathy (PVC-CM)—in which patients with idiopathic cardiomyopathy saw improvements in the left ventricular ejection fraction (LVEF) with pharmacologic suppression of PVCs.[3] Catheter ablation has also been successful in restoring ventricular function upon elimination of the PVCs.[4]

Numerous studies have revealed that the PVC burden correlates modestly with the extent of left ventricular dysfunction in attempts to risk-stratify which patients are likely to progress to PVC-CM.[5-7] In patients with a high PVC burden referred for ablation, the prevalence of PVC-CM is approximately 33%, with the lowest PVC burden resulting in PVC-CM being 10%.[7] However, some patients with a high PVC burden do not experience a drop in ejection fraction (EF) whereas others with a low burden progress to PVC-CM.[4,8,9] Historically, decisions to augment medical therapy or refer patients for catheter ablation are based on symptoms, PVC burden, and whether there is a reduction in LVEF.[7,10] In patients with a cardiomyopathy attributed to PVCs, 82% of patients undergoing catheter ablation procedures showed normalization of the LVEF within 6 months.[11] There is a dearth of clinical tools available to risk-stratify which patients are likely to experience a reduction in ventricular function in the presence of PVCs.[12]

FIGURE 1 Consort Diagram



ECG = electrocardiogram; LVEF = left ventricular ejection fraction; MI = myocardial infarction; PVC = premature ventricular contraction; TTE = transthoracic echocardiogram.

The ideal screening strategy would incorporate clinical information that is readily acquired in routine clinical practice.

Convolutional neural networks, which form the architectural basis for deep learning, a subset of machine learning and artificial intelligence, have frequently been used in image-based ECG analysis to make predictions and identify important features, which in turn have proven teachable to clinicians.[13,14] Although deep learning has been used to approximate the site of origin of PVCs based on the 12-lead ECG, there is a scarcity of data on its use to risk-stratify patients with PVCs who may stand to benefit from medical intervention.[5,15-17] Thus, there is a significant opportunity to apply deep learning to help clinicians identify which patients are at high risk of developing PVC-CM, with the future potential to consider changes in management based on the model prediction. To this end, we studied the possibility of using a deep learning algorithm to predict LVEF reduction in patients with PVCs based solely on a 12-lead ECG—an inexpensive, readily available, and frequently performed test.

TABLE 1 Patient Demographics (Testing and Validation)

	Total Cohort (N = 14,241; 27,370 ECGs)	EF >40% (n = 11,292; 20,941 ECGs)	EF ≤40 % (n = 3,262; 6,429 ECGs)
Age, y	67.6 ± 14.8	67.7 ± 15.0	67.6 ± 14.1
Male	56.2	52.7	68.2
Ethnicity			
White	29.5	30.9	25.3
Black	8.6	8.4	8.8
Hispanic	6.5	6.2	7.7
Asian	2.2	2.2	1.9
Other/unknown	53.3	52.2	56.3

Values are mean ± SD or %.
ECG = electrocardiogram; EF = ejection fraction.

METHODS

DATA SOURCES. We used available ECG data from 5 hospitals within the Mount Sinai Health System. These hospitals, namely, Mount Sinai Hospital, Mount Sinai Morningside, Mount Sinai West, Mount Sinai Beth Israel, and Mount Sinai Queens, serve a large demographically and socioeconomically diverse population in New York City. Data from the Mount Sinai Hospital were used for model training and testing, whereas pooled data from all other facilities were used for external validation.

ECG data were extracted from the GE MUSE system based on associated physician-confirmed diagnoses of PVCs. Values of LVEF, clinical notes, and International Classification of Diseases-10th Revision (ICD-10) codes were extracted from the electronic medical record and associated with ECGs based on unique patient identifiers. The institutional review board provided approval and ethical oversight for the conduction of the study.

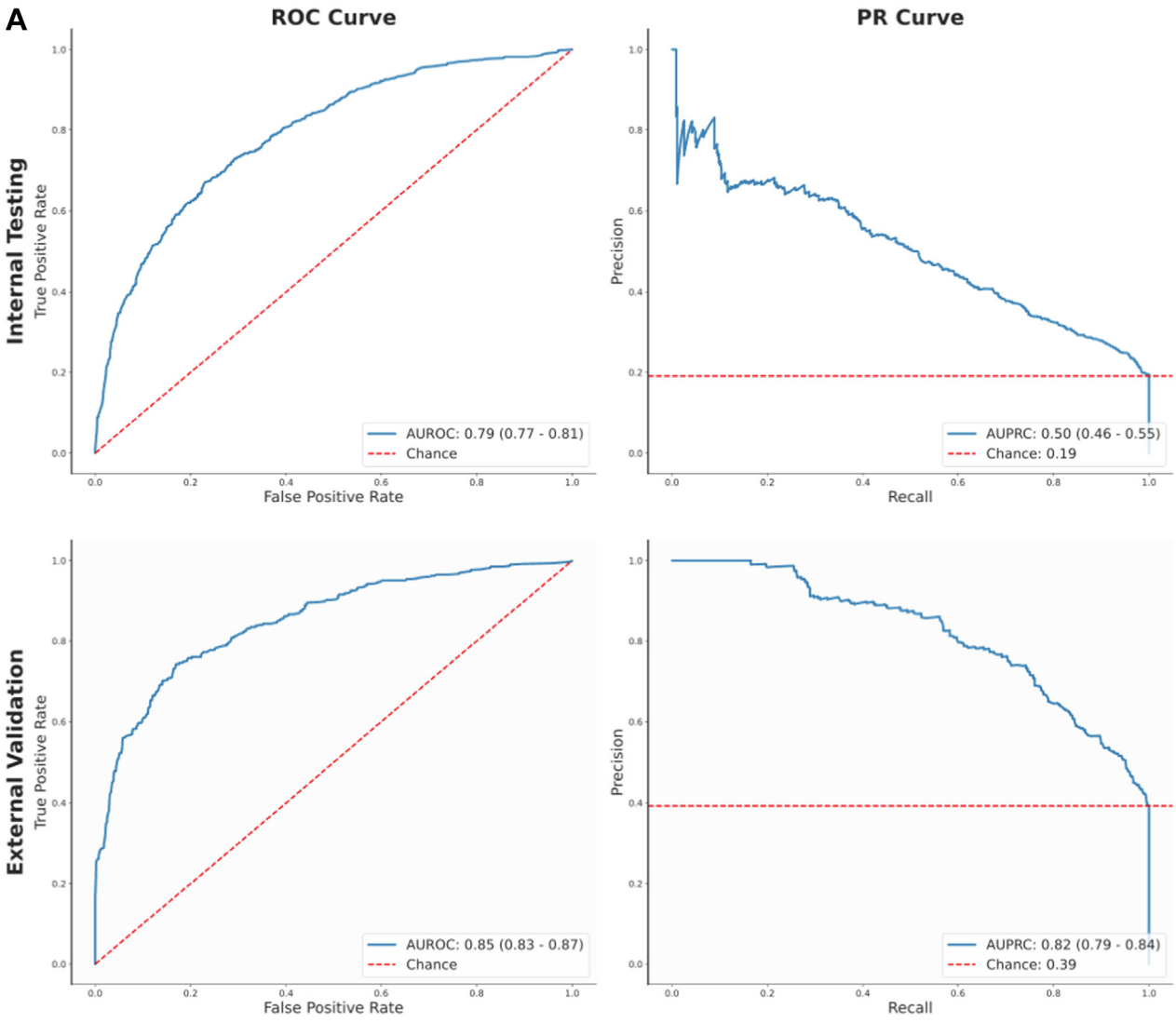
INCLUSION AND EXCLUSION CRITERIA. To be included, patients 18 years of age and older had to have serial ECGs of which at least 1 was required to exhibit physician-confirmed PVCs, an initial echocardiogram with a normal LVEF, and a follow-up echocardiogram. The time from the first ECG to the inclusion of the normal echo could not exceed 6 months. As patients had an ultimately normal echocardiogram at the end of that time frame, the LVEF was presumed normal for this time for predictive modeling. ECGs with a paced rhythm and ventricular tachycardia were excluded. Patients were excluded if they had an ICD-10 code associated with myocardial infarction or acute coronary syndrome at any time, even if not initially present at time of the initial echocardiogram.

PVC BURDEN AND ABLATION. To contextualize results, a manual review of patients who were flagged as positive by the algorithm were assessed for coincidental PVC ablation. A manual review was performed of patient-level data including notes, scanned monitor reports, the ablation procedural report, and (if applicable) device interrogation data. Drs Joshua Lampert and Vivek Reddy performed the manual review. There was no disagreement in interpretation or reporting of the outcomes data.

DATA PREPROCESSING. ECG data consists of XML files containing waveform data for leads I, II, and V_1-V_6. The remaining leads (III, aVF, aVL, and aVR) are called "derived leads" in that they only contain information present within other leads. ECG waveforms were subject to noise reduction by applying the Butterworth Bandpass filter, followed by a median filter. Resulting waveform data were plotted to images to allow use of 2-dimensional convolutional neural networks.

We created a rule-based natural language processing pipeline to parse clinical notes for PVC burden, which were obtained and calculated from ambulatory patch monitors and Holter monitor recordings. Regular expressions were created to extract any numbers expressed as percentages based on proximity to expressions of interest. Extracted values were again paired to ECGs based on unique patient identifiers.

FIGURE 2 Algorithm Performance



(A) Algorithm performance curves with ejection fraction (EF) cutoff ≤40%. **(B)** Algorithm performance EF <50%. AUPRC = area under precision recall curve; AUROC = area under the receiver operating characteristic curve; EF = ejection fraction; ROC = receiver operating characteristic.

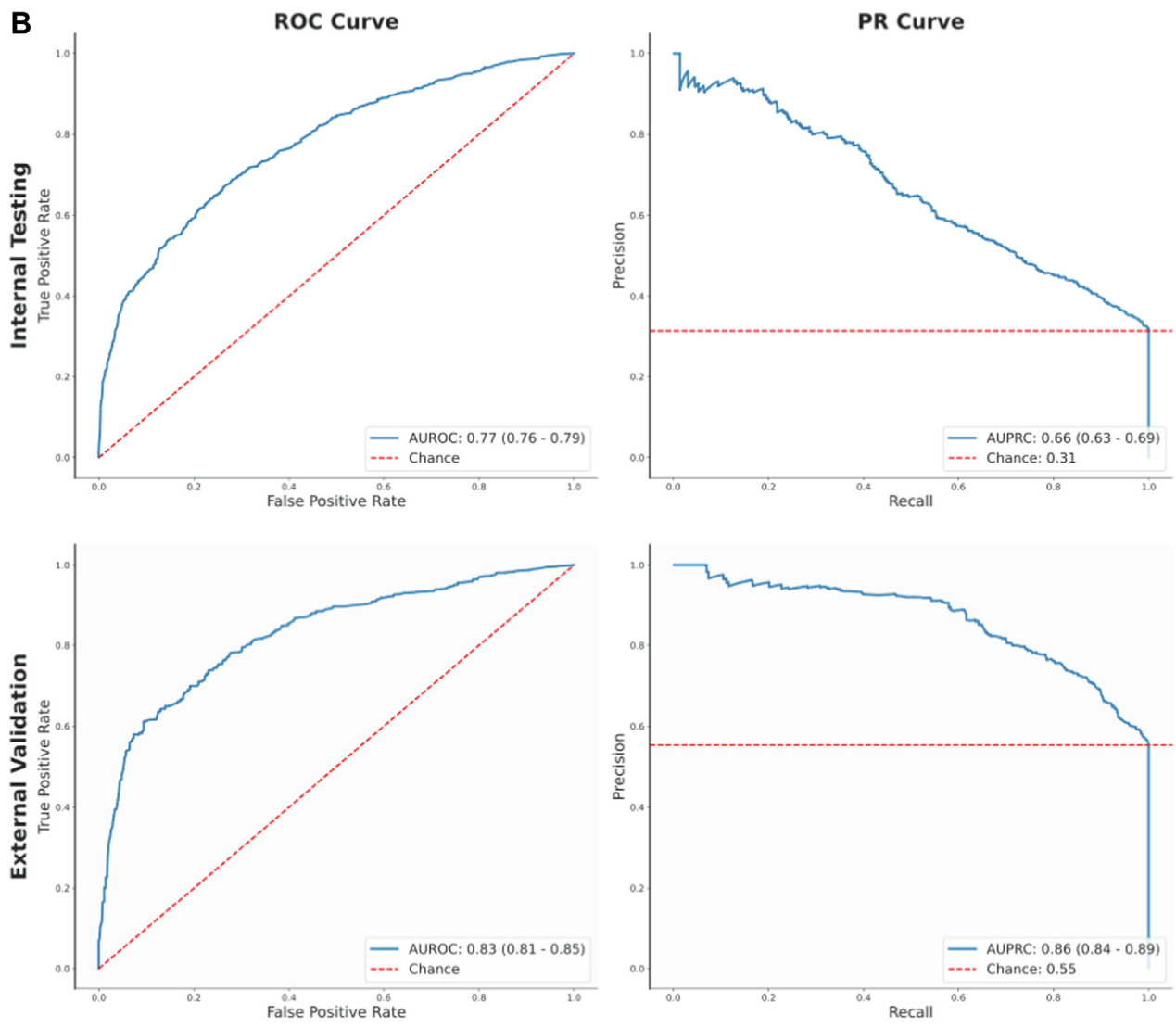
Continued on the next page

For patients with documented PVCs, we checked for the first instance of low LVEF as defined by a cutoff of either 40% or 50%. Patient data were discarded following a diagnosis of low LVEF, if they had an ICD-10-confirmed myocardial infarction, or if the first recorded LVEF value was below the cutoff.

DEFINITION OF PRIMARY OUTCOME. An ECG was labeled as positive for the outcome in case a patient with PVCs developed cardiomyopathy as defined by a fall in the LVEF to below the cutoff within 6 months of the date of the ECG. Because there are only 2 possible states for the outcome variable, the task may be considered a binary classification problem.

MODEL DEVELOPMENT AND EVALUATION. We selected the largest available pretrained ResNet model (ResNet-152) as the starting point for our analyses. Using models pretrained on natural images allows for better performance with less data, while also requiring less time to achieve an optimal solution.

FIGURE 2 **Continued**

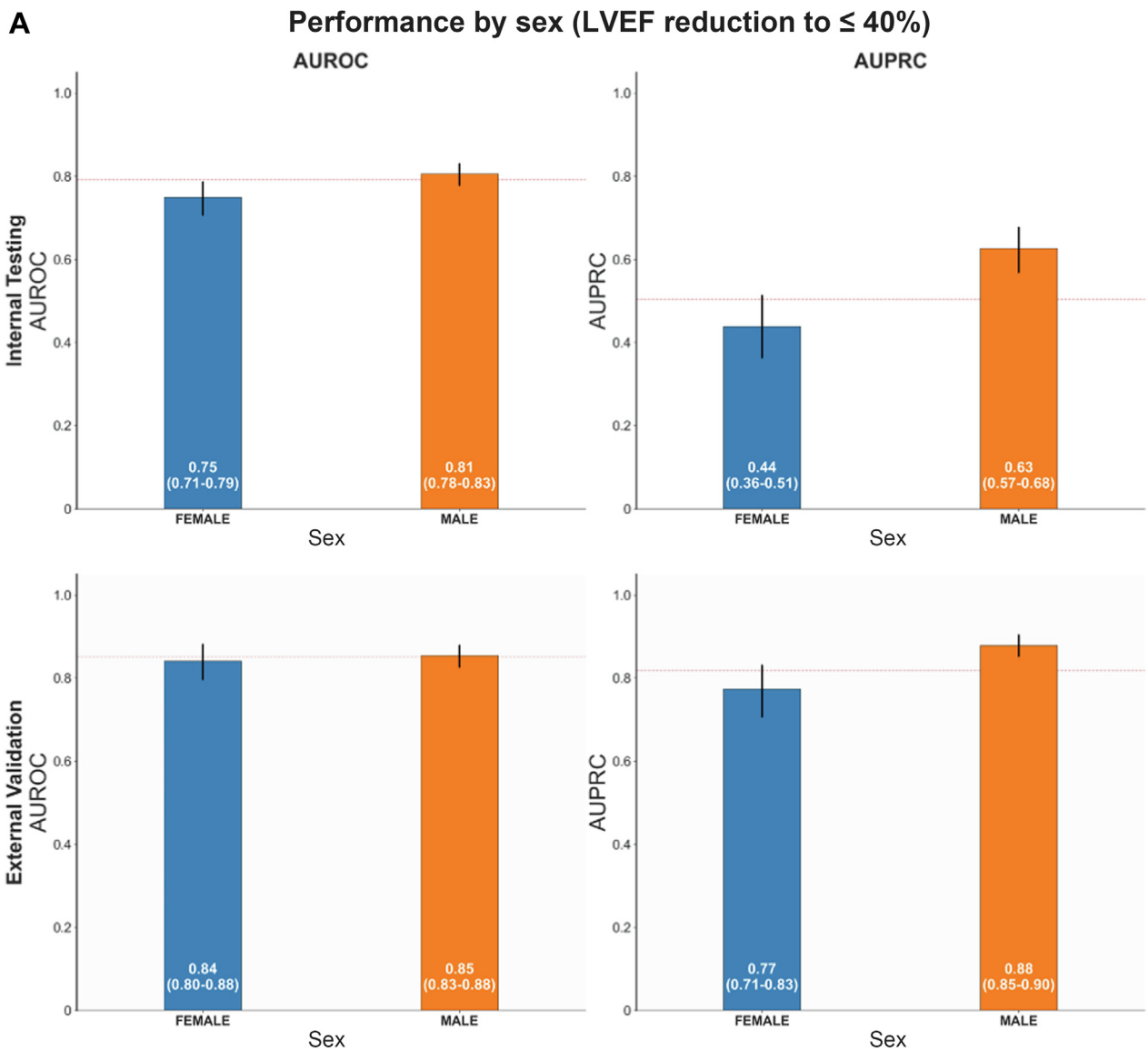


Data were split based on group shuffle splitting which removes the potential for data leakage by ensuring that no patients are present in both the training and testing groups. We elected to use the Adam optimizer with a learning rate of 3e-4 on a OneCycle learning rate schedule. Models were trained for 35 epochs, and performance reported on the epoch with the best performance on internal testing data. Resulting models were also separately evaluated on external validation data. To avoid overfitting, we used learning rate scheduling alongside continual logging of loss and performance metrics to achieve the best fit for the model. Training was ceased when model performance started to deteriorate (overfitting occurred) after initial improvement. The internal cohort refers to the patients from Mount Sinai Hospital whereas the external cohort were comprised of the external validation dataset obtained from pooled patients from the other 4 affiliated hospitals.

We used the area under the receiver operating characteristic curve (AUROC), and area under the precision recall curve (AUPRC) metrics to evaluate model performance. These metrics use the uncalibrated probability estimates output from the model to generate curves which signify the model's ability to discriminate positives from negatives.

Calibrated Cox proportional hazard models were generated for all patients with PVC burdens identified

FIGURE 3 Model Performance by Male or Female Sex



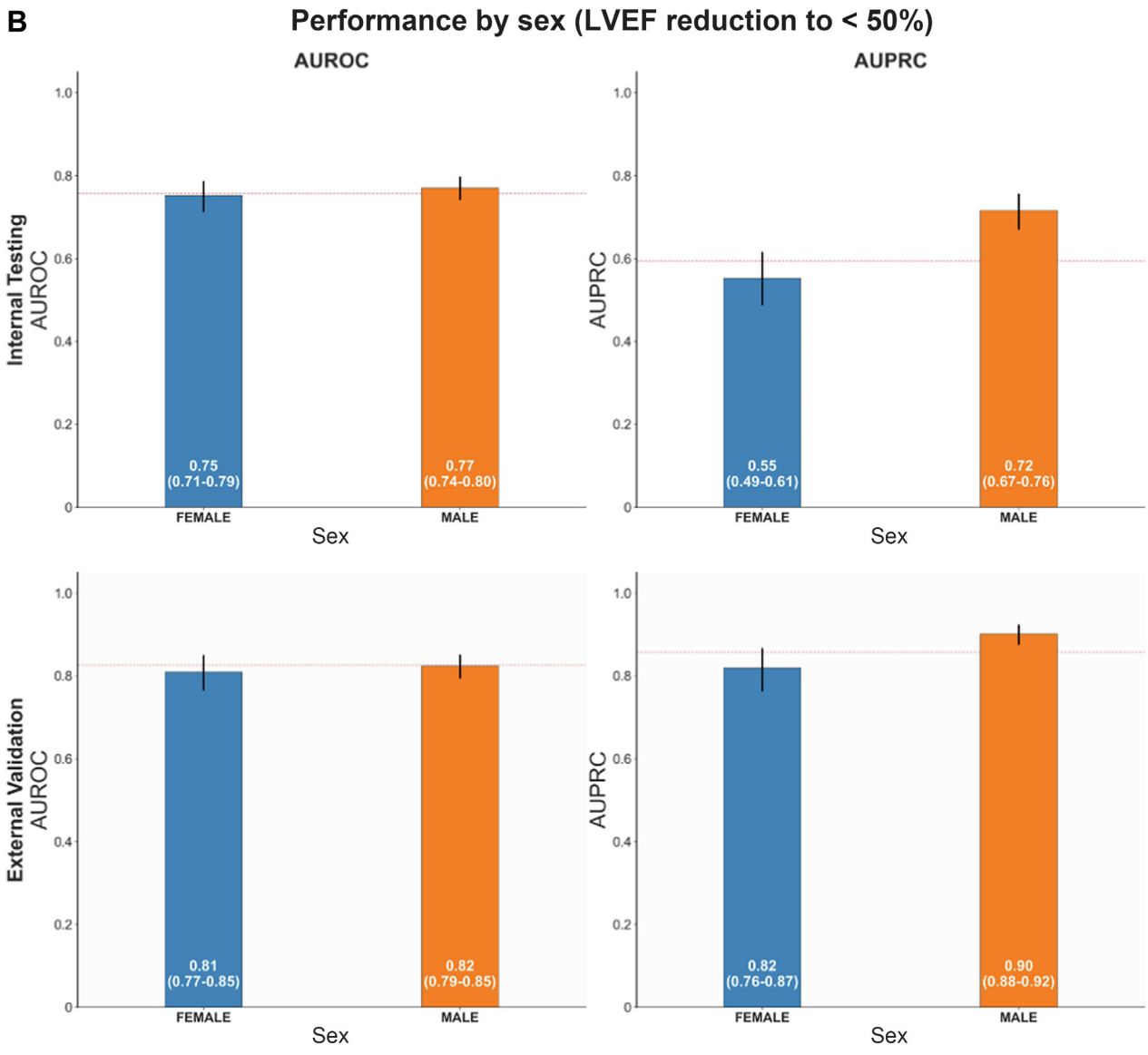
(A) Performance by sex with EF cutoff ≤40%. **Red line** denotes overall performance (AUROC). **(B)** Performance by sex with EF cutoff <50%. AUPRC = area under the precision recall curve; other abbreviation as in Figure 2.

Continued on the next page

by natural language processing to assess model performance in context of age, sex, and PVC burden as covariates. For the purposes of these models, a positive model output (artificial intelligence-positive [AI+]) was defined as a ≥80% probabilistic model output for developing a reduced ejection fraction to <50% with 80% specificity. We performed calibration using isotonic regression on the output of the neural network .

EXPLAINABILITY. We used gradient-weighted class activation mapping (GradCAM) methodology for generating class activation maps. These maps show which areas of the ECG are the most responsible for pushing the model towards a prediction. To further assess algorithm prediction, we extracted patients for whom the model predicated subsequent LVEF impairment with ≥80% probability with 80% specificity who had ablation by checking for acute

FIGURE 3 Continued



blood loss anemia as a wildcard. Then, a manual chart review was performed by an electrophysiologist. We excluded patients if they did not have follow-up echocardiographic data within 2 years of ablation. We determined presence of PVCs at follow-up by acquired ECGs, monitors, and device interrogations including presenting electrograms. We used a paired, 2-tailed Student's *t*-test to assess changes in EF in patients who underwent PVC ablation.

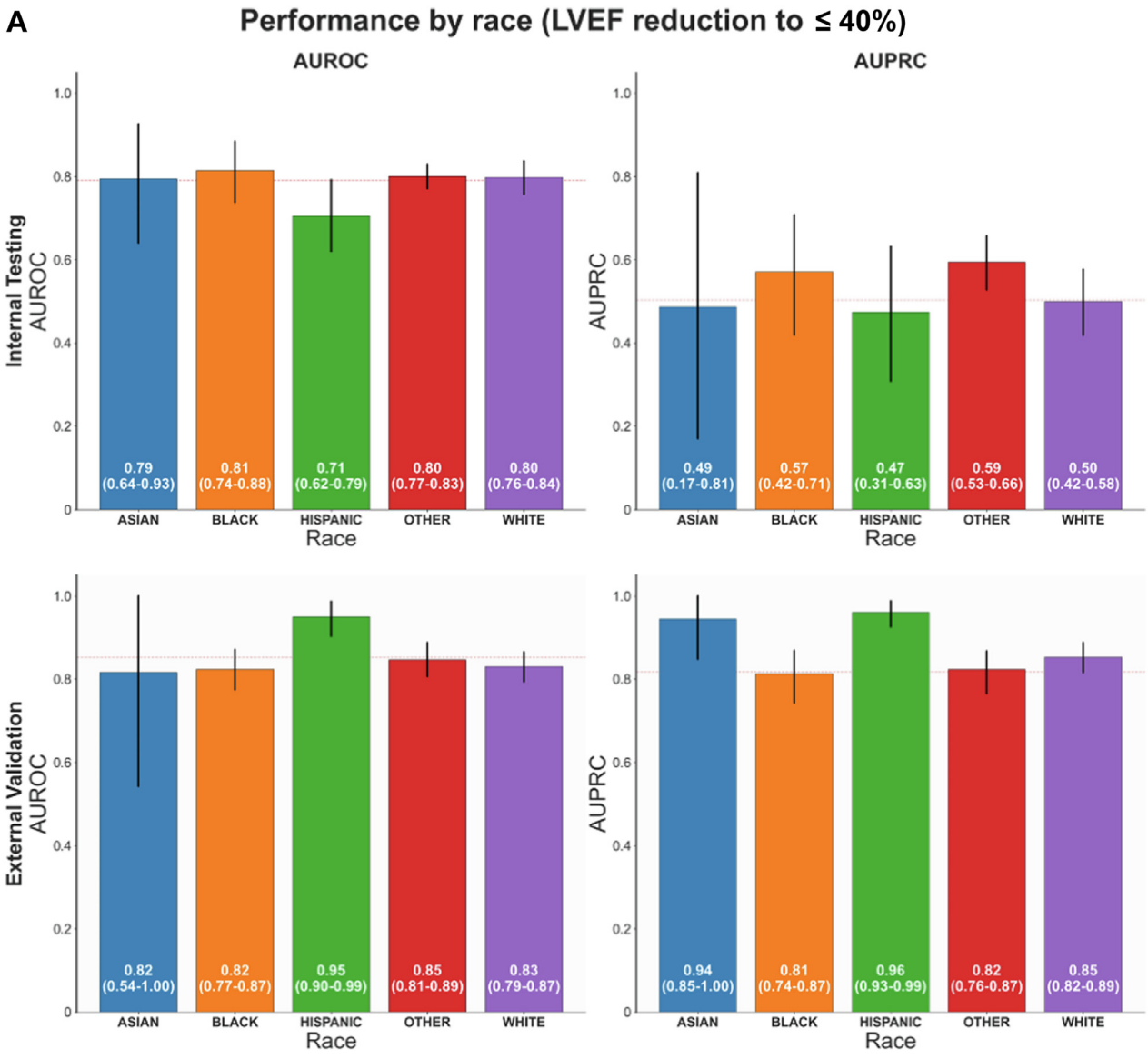
SOFTWARE AND HARDWARE. We performed analyses using the numpy, pandas, scipy, scikit-learn, PyTorch, and torchvision libraries inside a custom virtual environment. We performed plotting using the matplotlib and seaborn libraries. All code was written for and within the Python programming language (3.8.x).

CODE AVAILABILITY. The code will be made available upon publication.

RESULTS

STUDY POPULATION. The algorithm was trained and tested using 13,553 patients with 25,995 ECGs from

FIGURE 4 Model Performance by Race



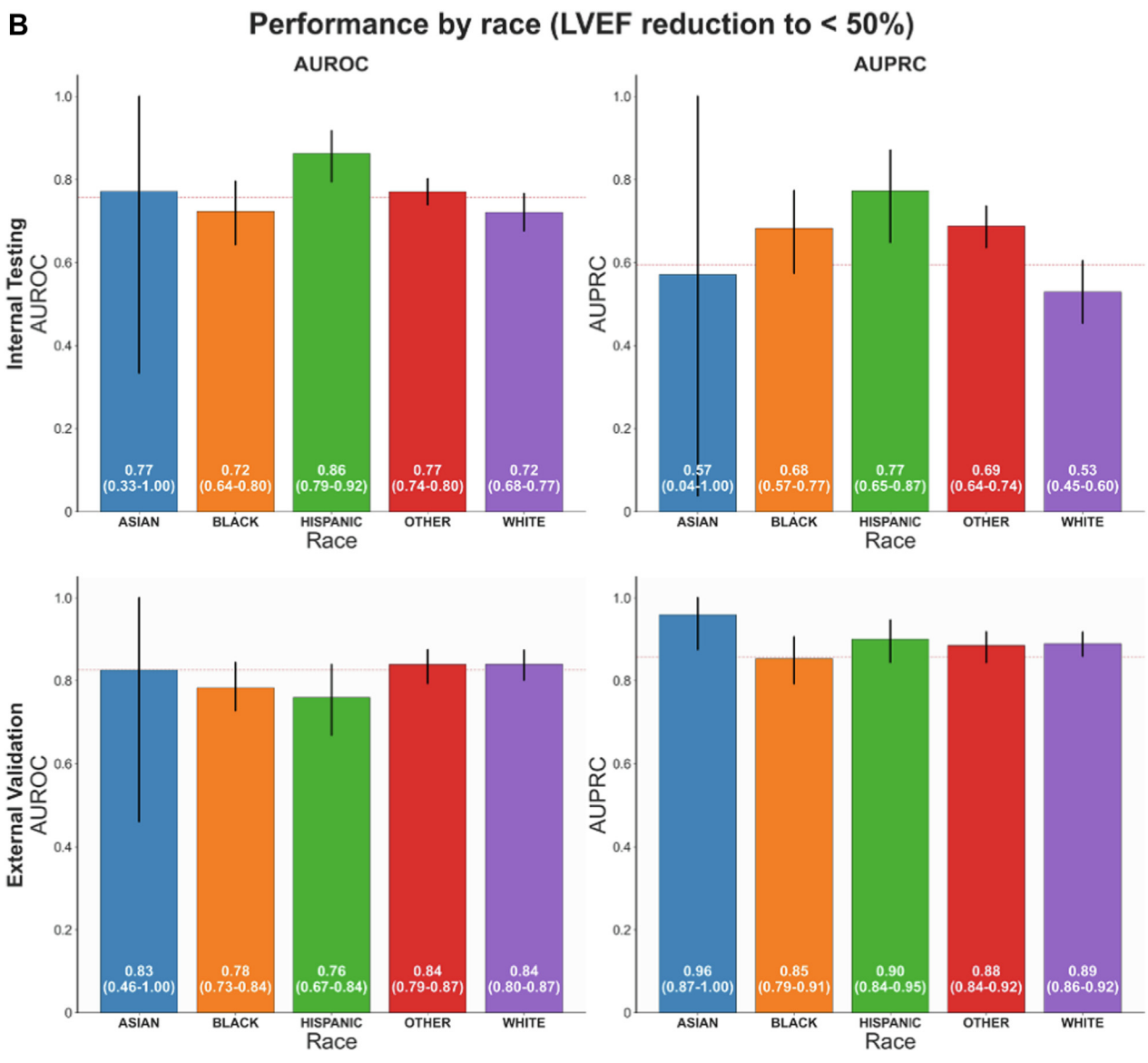
(A) Performance by race for EF cutoff ≤40%. **(B)** Performance by race for EF cutoff <50%. **Red line** denotes overall performance (AUROC). Abbreviations as in Figures 1-3.

Continued on the next page

one hospital (80% were used for training and 20% for testing with group shuffle splitting), and externally validated with pooled data from the other 4 hospitals (n = 688 and 1,375 ECGs) (Figure 1). Patient demographics are summarized in Table 1. The mean age was 67.6 ± 14.8 years. Patient races included 29.5% White, 8.6% Black, 6.5% Hispanic, 2.2% Asian, and 53.3% other or unknown race. The cohort was 56.2% male.

The mean age of patients who experienced a reduction in EF to ≤40% or <50% were 67.6, and 67.6 ±14.7, respectively. The total prevalence of a follow-up LVEF of ≤40% or <50% was 22.9% or 39%, respectively. The prevalence of a reduced EF to ≤40% was 19% in the internal cohort and 39% in the external cohort. As expected, for a reduction to LVEF <50%, the prevalence was higher in the internal cohort (31%) and external cohorts (55%).

FIGURE 4 Continued



MODEL PERFORMANCE: INTERNAL TESTING. The model was trained on the 12-lead ECG to detect a reduction in EF to ≤40% and <50% by echocardiogram at 6 months. Using a follow-up LVEF cutoff of ≤40%, the model performed well with an AUROC of 0.79 (95% CI: 0.77-0.81). The AUPRC was 0.50 (95% CI: 0.46-0.55) (Figure 1). As shown in Figure 2, when an LVEF cutoff of <50% was used, the AUROC and AUPRC were 0.77 (95% CI: 0.76-0.79) and 0.66 (95% CI: 0.63-0.69), respectively.

The AUROC curve was generated to assess model performance across differences in patient sex and ethnicity. When assessing an EF cutoff of ≤40%, the AUROC as shown in Figure 3A was 0.81 (95% CI: 0.71-0.79, AUPRC 0.63) in men and 0.75 in women (95% CI; 0.71-0.79, AUPRC 0.44). As shown in Figure 4A, the AUROC was 0.80 (95% CI: 0.76-0.84) in White patients, 0.81 (95% CI: 0.74-0.88) in Black patients, 0.71 (95% CI: 0.62-0.79) in Hispanic patients, and 0.79 (95% CI: 0.64-0.93) in Asian patients. Comparable performance was observed when assessing for an EF cutoff of <50%: for sex, the AUROC was 0.77 (95% CI: 0.74-0.80, AUPRC 0.72) in men and 0.75 (95% CI :0.71-0.79,AUPRC 0.55) in women (Figure 3B), and for ethnicity—White, Black, Hispanic, and Asian—the AUROCs were 0.72 (95% CI: 0.68-0.77), 0.72 (95% CI:

TABLE 2 Multivariable Cox Regression Model for the Outcome of LVEF <50%

	HR	95% CI	*P* Value
Algorithm predicting cardiomyopathy	333.6	122.3-909.8	<0.01
PVC burden	1.32	1.2-1.5	<0.01
Age	0.78	0.71-0.87	<0.01
Male	1.4	1.1-1.8	0.01

The algorithm covariate is a categorical variable with a positive algorithm prediction of subsequent left ventricular ejection fraction (LVEF) impairment with a ≥80% probabilistic output at 80% specificity. Premature ventricular contraction (PVC) burden and age are continuous variables. Continuous variables were standardized using z-score standardization.

TABLE 3 Multivariable Cox Regression Model for the Outcome of LVEF ≤40%

	HR	95% CI	*P* Value
Algorithm predicting cardiomyopathy	304.1	77.2-1,197.5	<0.01
PVC burden	1.42	1.3-1.6	<0.01
Age	0.78	0.71-0.85	<0.01
Male	1.6	1.2-2.0	<0.01

The algorithm covariate is a categorical variable with a positive algorithm prediction of subsequent LVEF impairment with a ≥80% probabilistic output at 80% specificity. Premature ventricular contraction PVC burden and age are continuous variables. Continuous variables were standardized using z-score standardization.

Abbreviations as in Table 2.

0.64-0.80), 0.86 (95% CI: 0.79-0.92), and 0.77 (95% CI: 0.33-1.00), respectively (Figure 4B).

EXTERNAL VALIDATION. We additionally evaluated algorithm performance in the pooled and aggregated cohort from across the 4 external hospitals within the health care system. The model performed well with AUROC of 0.85 (95% CI: 0.83-0.87) and AUPRC of 0.82 (95% CI: 0.79-0.84) using an EF cutoff of ≤40%. The model performed well in both male (AUROC 0.85, 95% CI: 0.83-0.88, AUPRC 0.88) and female (AUROC 0.84, 95% CI: 0.80-0.88, AUPRC 0.77) patients (Figure 3A). Across races, the algorithm also performed well. The AUROCs were 0.83 (95% CI: 0.79-0.87), 0.82 (95% CI: 0.77-0.87), 0.95 (95% CI: 0.90-0.99), and 0.82 (95% CI: 0.54-1.00), among White, Black, Hispanic, and Asian patients, respectively (Figure 4A).

When considering development of a reduced EF to <50%, the AUROC was 0.83 (95% CI: 0.81-0.85) and the AUPRC was 0.86 (95% CI: 0.84-0.89). Model performance was comparable across men (AUROC 0.82, 95% CI: 0.79-0.85, AUPRC 0.90) and women (AUROC 0.81, 95% CI: 0.77-0.85, AUPRC 0.82) as shown in Figure 3B. Likewise, the model performed well across differences in race on external validation. The AUROCs were 0.84 (95% CI: 0.80-0.87), 0.78 (95% CI: 0.73-0.84), 0.76 (95% CI: 0.67-0.84), and 0.83 (95% CI: 0.46-1.00) among White, Black, Hispanic, and Asian patients, respectively (Figure 4B).

PVC BURDEN. Derived from the natural language processing pipeline of patients with a positive algorithm prediction (AI+) and a PVC burden, 2,517 ECGs remained for the calibrated model evaluating a cutoff of LVEF ≤40% after 2,129 right-censored ECGs were excluded. A total of 2,006 paired ECGs remained for analysis with an LVEF cutoff of <50% after exclusion of 1,700 right-censored ECGs paired to the outcome. Positive algorithm prediction (AI+) was a strong negative independent prognostic factor. As shown in Table 2, using an EF cutoff of <50%, a positive algorithm prediction was associated with a more than 300-fold increased risk in subsequent LVEF impairment (HR: 333.6; 95% CI: 122.3-909.82; $P < 0.01$). This association was independent of PVC burden in the Cox model, whereas PVC burden was also an independent, albeit weaker, predictor of a reduction in EF (HR: 1.32; 95% CI: 1.2-1.5; $P = 0.01$). For an LVEF cutoff of ≤40%, AI+ remained a strong independent predictor of cardiomyopathy (HR: 304.1; 95% CI: 77.2-1197.5; $P < 0.01$). Although AI+ portended vastly increased likelihood of subsequent systolic functional impairment independent of PVC burden, the actual PVC burden (as a continuous variable) was also an independent predictor of cardiomyopathy in this group (HR: 1.42; 95% CI: 1.3-1.6; $P < 0.01$) (Table 3). Model classification metrics are summarized in Table 4. When assessing an EF cutoff of ≤40%, the model had a sensitivity of 62% and specificity of 80% with a positive predictive value (PPV) of 42% and negative predictive value of 90% on internal testing. The PPV was higher (69%) on external testing. Similarly, the PPV was 55% on internal testing with an EF cutoff of <50% but was 81% when model prediction was applied to the external validation cohort. The highest model sensitivity was observed in the external cohort with an EF cutoff of ≤40% at 77%.

ASSESSMENT OF AI+ PATIENTS WHO UNDERWENT PVC ABLATION. To determine whether the model had identified patients where PVCs directly caused cardiomyopathy, as opposed to an unrelated correlation, we determined the effect of coincident PVC ablation on outcomes by manual retrospective chart review. Of the 286 AI+ patients who underwent any ablation procedure, the procedure was performed to target PVCs in 35 patients; among this subset, 15 patients also had subsequent follow-up echocardiographic data for review.

TABLE 4 Model Classification Metrics

EF Cutoff	Dataset	Sensitivity	Specificity	PPV	NPV
≤40%	Internal	0.62	0.80	0.42	0.90
≤40%	External	0.77	0.77	0.69	0.84
<50%	Internal	0.56	0.80	0.55	0.81
<50%	External	0.66	0.83	0.81	0.69

Internal dataset includes the internal testing cohort. External dataset includes the external validation cohort. Model positivity was deemed positive at the respective EF cutoff with ≥80% model probability prediction.

NPV = negative predictive value; PPV = positive predictive value; other abbreviation as in Table 1.

TABLE 5 Impact of PVC Ablation in Retrospective Cohort

	Pre-Ablation LVEF, %	Post-Ablation LVEF, %	*P* Value
Full	32.9	45.9	<0.01
LVEF improvement	34.4	54.7	<0.01
No LVEF improvement	30.7	32.7	0.08

Improvement in LVEF post-ablation was defined by an increase in LVEF by ≥10% or improvement to LVEF >50%.

Abbreviation as in Table 2.

Following the ablation procedure, 9 of 15 (60%) patients had an improvement in EF after follow-up. The LVEF improved from a mean of 34.4% to 54.7% (Table 5), with 89% having a follow-up EF exceeding 50%. Of the 6 patients who did not show LVEF improvement (pre- and post-LVEFs being 30.7% and 32.7%, respectively), all patients (100%) were found to have recurrent PVCs at follow up—that is, unsuccessful PVC ablation procedures.

EXPLAINABILITY FRAMEWORK. GradCAM was performed to assess which features were important in driving the model to an outcome prediction. Model prediction focused primarily on lead II and lead I. A heat map of feature importance is presented in Figure 5 showing that the algorithm highlighted the QRS complex and ST segment in sinus rhythm to be the most important in predicting LVEF impairment. PVCs were not highlighted. Additionally, QRS complex and ST-segment during ectopy, such as premature atrial contractions as shown in Figure 5D, were also not highlighted as important features.

DISCUSSION

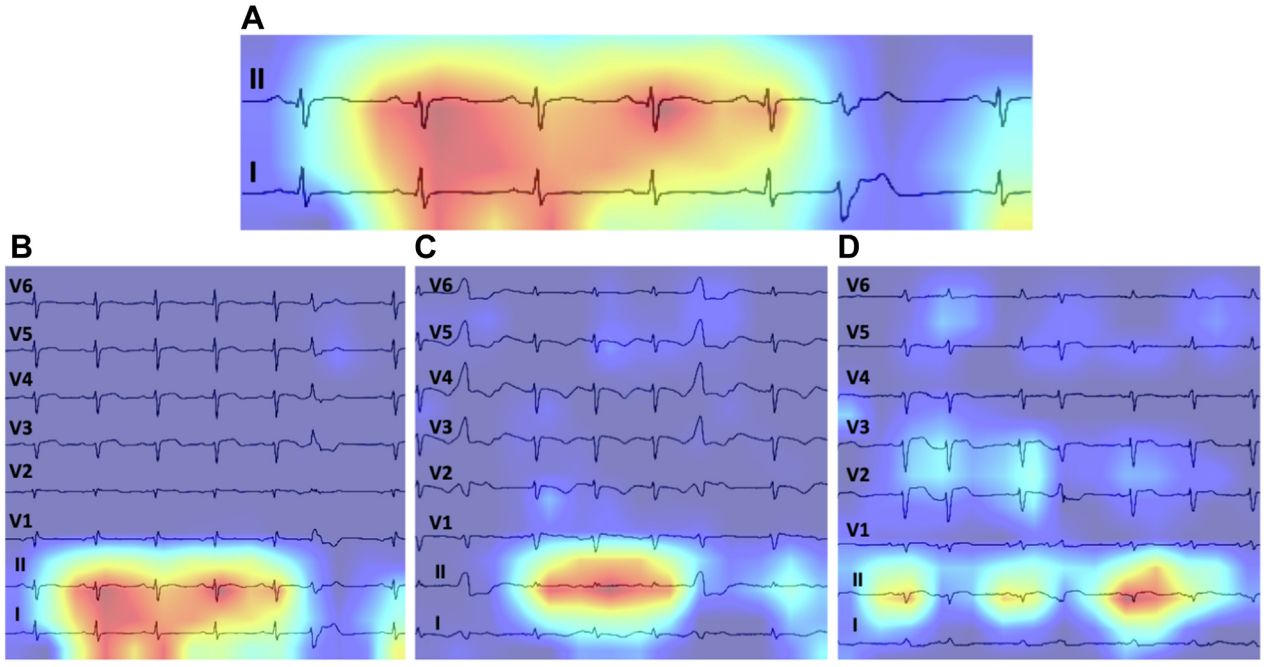
This study presents a novel deep learning algorithm based on the 12-lead ECG for predicting a reduction in LVEF to ≤40% and <50% within 6 months in patients with PVCs and without a history of myocardial infarction. The most important findings of our study are as follows: 1) the algorithm can accurately predict cardiomyopathy within 6 months; 2) the algorithm was accurate for either sex, and across a range of ethnicities; 3) algorithm prediction is independent of PVC burden; 4) the algorithm was a compelling and independent predictor of a reduction in EF over a finite, actionable clinical period; 5) the explainability analysis highlighted the importance of the sinus beat QRS complex and ST-segment, and not the PVC morphology itself, suggesting that PVC origin is not driving cardiomyopathy development and that model prediction may be identifying vulnerable myocardium at risk for decompensation when exposed to a stressor such as PVCs; and 6) the LVEF improved in a subset of patients who happened to undergo successful PVC ablation procedures.

Perhaps the most intriguing observation in this study is that positive algorithm prediction is not only associated with vastly increased likelihood of subsequent left ventricular systolic function impairment, but it is also independent of PVC burden. This suggests that frequency of PVCs, a cornerstone metric of current clinical decision making, is not the primary driver of cardiomyopathy development in context of algorithm prediction in this patient population.[12]

Another notable aspect of this study was the use of explainable AI to highlight relevant portions of the ECG.[18] GradCAM is a technique that allows visualization of the feature(s) of the input data, in this case the ECG waveform, that are most important in model prediction. In our study, GradCAM highlighted the QRS complex and ST-segment in sinus rhythm as the most important features contributing to model prediction, particularly in lead II (Figure 5). This finding was somewhat unexpected given historical risk factors associated with PVC-CM including high PVC burden (>10% to 15%), nonsustained ventricular tachycardia, retrograde p-waves following the PVC, QRS duration during PVC ≥140 milliseconds, PVC morphology or origin, and PVC coupling interval or interpolation.[12,19-23]

GradCAM attribution of feature importance, shown in Figure 5, not only focuses on the sinus QRS complex in our study, but also potentially builds on the recent work by Alhede et al[24] that showed mechanically abnormal sinus beats, as defined by echocardiographic global longitudinal strain imaging, before, during, and after PVCs. The mechanisms postulated in that paper include abnormal autonomic inputs and excitation-contraction coupling; however, no discrete, single mechanism was implicated. The mechanically abnormal sinus beat presaging the PVC in that paper is consistent with our explainability

FIGURE 5 GradCAM Attributions of Feature Importance



(A) Example of gradient-weighted class activation mapping (GradCAM) attribution of feature importance. Across patients, lead II was highlighted most frequently in pushing the model towards a prediction. **(B)** Patient example with GradCAM highlighting sinus QRS and ST-segment as important features. **(C)** Patient example with GradCAM highlighting sinus rhythm QRS complexes, focusing on lead II. PVCs are not highlighted despite 25% premature ventricular contraction (PVC) burden on this recording. **(D)** Patient example with GradCAM highlighting sinus rhythm QRS complexes. The PVC is not highlighted.

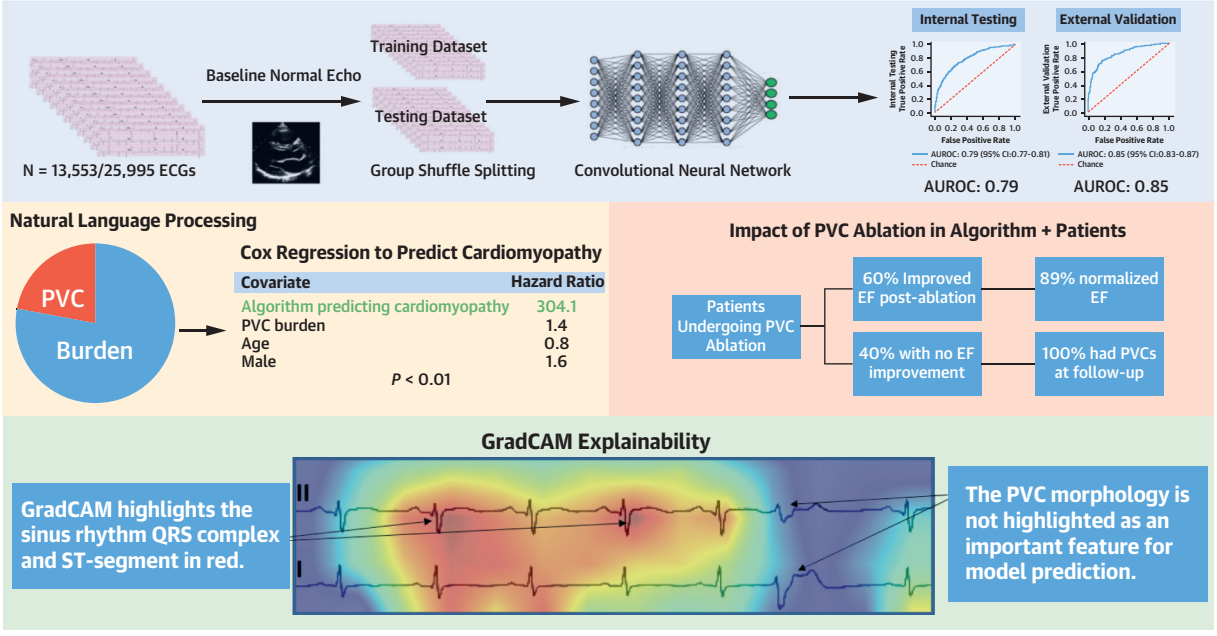
model—together suggesting that PVC morphology, and therefore origin, is likely not the most important feature mechanistically to explain the development of a cardiomyopathy.

Likewise, our finding that the QRS complex in sinus rhythm is a predictive feature of subsequent cardiomyopathy development in patients to the exclusion of PVC morphology and coupling interval challenges previous work highlighting the importance of metrics based solely on PVC parameters such as duration and whether retrograde conduction was present.[12,19,20] Alternatively, the importance emphasized in our study of depolarization in sinus rhythm is consistent with previous work implicating aberrant excitation-contraction coupling, calcium metabolism, and/or autonomic dysregulation in cardiomyopathy development, although we cannot strongly implicate a single mechanism.[4,25-27]

In our manual review of AI+ patient charts, 60% of patients analyzed improved after PVC ablation and 89% of those patients experienced essentially full recovery of systolic function, supporting that these AI+ patients truly had a PVC-CM. Two findings from this analysis are notable. First, all patients who did not improve after ablation were noted to have PVCs at follow-up. Second, there is a large discrepancy between the number of patients identified by the algorithm as being at high risk for developing a cardiomyopathy and the number who underwent ablation for PVCs. The latter finding suggests significant potential for the algorithm to meaningfully identify patients at risk of developing cardiomyopathy who may benefit from augmented care.

One finding in this analysis is that the area under the curve (AUC) was higher in the external validation cohort. There are several plausible explanations for this. First, in context, the internal cohort was diverse and model overfitting was unlikely to contribute to an inappropriately high AUC in the cohort to which the comparison is made. Second, the external validation was performed with pooled data from 4 external hospitals within the health system comprising diverse patient populations with variable prevalence of disease (which can also describe the increased AUPRC). In this circumstance, 2 specific scenarios can explain a higher AUC. The first is that the standard deviation of model performance can vary when the model is applied to the external cohort. However, a

CENTRAL ILLUSTRATION **Deep Learning to Predict Cardiomyopathy in Patients With PVCs**



Lampert J, et al. J Am Coll Cardiol EP. 2023;9(8):1437–1451.

AUROC = area under the receiver operating characteristic curve; ECG = electrocardiogram; EF = ejection fraction; GradCAM = gradient weighted class activation map; LV = left ventricular; PVC = premature ventricular contraction.

more likely contributor in these data could be a function of increased separation in correlation variance of cases in the external cohort. For example, if negative cases were deemed to be very clearly negative (highly negative correlation), and positive cases were very positive (highly positive correlation) at opposite ends of the AUROC curve, or with a higher case mix variability, the resulting AUC would increase. To phrase differently, the AUC can vary purely as variance in correlation of cases (patients who experience the outcome) even when correlation of controls is fixed at different levels. This again fits with differences in disease prevalence between the cohorts as consistent patterns of AUC change are not observed if the effect size is small. Lastly, in this retrospective study, confounding could also contribute to such variance.

Currently, this novel deep learning algorithm may be clinically helpful via incorporation in shared decision-making with patients with PVCs when discussing management options such as pursuing diagnostic imaging, medical therapy, or ablation in appropriate patients.[28,29] This is particularly intriguing because the data was based on a single 12-lead ECG—a relatively ubiquitous test conducted in most medical practices. However, caution is advised particularly when considering invasive therapeutic options based on this model prediction until prospective studies confirm these results. Ultimately, this algorithm shows promise for meaningful incorporation into clinical risk stratification given the ubiquitous and inexpensive ECG basis for analysis.

STUDY LIMITATIONS. This retrospective study has several limitations. First, although we excluded patients with a history of or acute myocardial infarction, which should limit contribution of an ischemia to the development of a cardiomyopathy, our methodology has not completely excluded other potential etiologies such as tachycardia-induced cardiomyopathy, stress-induced cardiomyopathy, or myocarditis. Although such additional diagnoses could potentially explain cardiomyopathy development were not accounted for, the timeframe of a reduction in EF within 6 months limits the differential and mitigates confounding due to other diagnoses that could explain reductions in EF. Likewise, the prevalence of disease noted in our study is not consistent with reported prevalence of these other diagnoses that could result in a cardiomyopathy over a 6-month window.[30-35] Furthermore, we relied on extracted XML

files for screening ECG data which could have resulted in some mislabeled ECGs. To mitigate this risk, we manually reviewed representative sets of ECGs of each diagnostic code to ensure we identified patients with true PVCs. Likewise, although the manual electrophysiologist review of patient-level data bolsters data quality, the analysis is limited by the small sample size. On the other hand, the systematic fashion by which this analysis was performed mitigates the possibility of confounding issues. Lastly, patients were analyzed retrospectively without standardized acquisition intervals between ECG and echocardiography. A prospective cohort with standardized imaging intervals could better obviate selection bias.

CONCLUSIONS

Deep learning applied to the 12-lead ECG in patients with PVCs can predict a reduction in LVEF whether a cutoff value of ≤40% or <50% is used in model prediction. Algorithm prediction is associated with vastly increased adjusted likelihood of subsequent systolic dysfunction independent of PVC burden. The algorithm performed well both on external validation as well as across differences in sex and race. Whereas the QRS complex in sinus rhythm appears to be important in predicting a decrease in LVEF, PVC morphology does not appear to be an important feature driving model prediction, suggesting model prediction may identify vulnerable myocardium at higher risk for cardiomyopathy when exposed to a stressor such as PVCs. Further studies should assess whether incorporation of this algorithm into clinical decision-making would change both quality metrics and clinical outcomes.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Dr Lampert has served as a consultant for Viz.AI. Dr Reddy has served as a consultant for and has equity in Ablacon, Acutus Medical, Affera-Medtronic, Anumana, Apama Medical-Boston Scientific, APN Health, Aquaheart, Atacor, Autonomix, Axon Therapies, Backbeat, BioSig, CardiaCare, CardioNXT/AFTx, Circa Scientific, CoRISMA, Corvia Medical, Dinova-Hangzhou DiNovA EP Technology, East End Medical, EPD-Philips, EP Frontiers, Epix Therapeutics-Medtronic, EpiEP, Eximo, Farapulse-Boston Scientific, Field Medical, Focused Therapeutics, HRT, Intershunt, Javelin, Kardium, Keystone Heart, LuxMed, Medlumics, Middlepeak, Neutrace, Nuvera-Biosense Webster, Oracle Health, Restore Medical, Sirona Medical, SoundCath, Valcare unrelated to this work; has served as a consultant for Abbott, AtriAN, Biosense-Webster, BioTel Heart, Biotronik, Boston Scientific, Cairdac, Cardiofocus, Cardionomic, CoreMap, Fire1, Gore & Associates, Impulse Dynamics, Medtronic, Novartis, Philips, and Pulse Biosciences; and has equity in Manual Surgical Sciences, Newpace, Nyra Medical, Surecor, and Vizaramed. Dr Nadkarni reports consultancy agreements with AstraZeneca, BioVie, GLG Consulting, Pensieve Health, Reata, Renalytix, Siemens Healthineers and Variant Bio; has received research funding from Goldfinch Bio, and Renalytix; has received honoraria from AstraZeneca, BioVie, Lexicon, Daiichi Sankyo, Meanrini Health, and Reata; has patents or royalties with Renalytix; owns equity and stock options in Pensieve Health and Renalytix as a scientific cofounder; owns equity in Verici Dx; has received financial compensation as a scientific board member and advisor to Renalytix; has served on the advisory board of Neurona Health; and has served in an advisory or leadership role for Pensieve Health and Renalytix. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Vivek Y. Reddy, Helmsley Electrophysiology Center, Mount Sinai Hospital Electrophysiology, 1190 Fifth Avenue, GP 1 South, New York, New York 10029, USA. E-mail: vivek.reddy@mountsinai.org.

PERSPECTIVES

COMPETENCY IN PATIENT CARE AND PROCEDURAL SKILLS: A novel deep learning algorithm based solely on a 12-lead ECG can predict the development of LVEF impairment in patients with PVCs, and a GradCAM explainability framework suggests that the most important feature contributing to the model prediction is the sinus beat QRS complex and ST-segment.

TRANSLATIONAL OUTLOOK: Further studies are warranted to assess how the incorporation of a deep learning algorithm into clinical practice to predict the onset of cardiomyopathy in patients with PVCs may impact clinical workflow, treatment decision-making, and patient outcomes.

REFERENCES

1. Kennedy HL, Whitlock JA, Sprague MK, Kennedy LJ, Buckingham TA, Goldberg RJ. Long-term follow-up of asymptomatic healthy subjects with frequent and complex ventricular ectopy. *N Engl J Med*. 1985;312(4):193–197.

2. Ng GA. Treating patients with ventricular ectopic beats. *Heart*. 2006;92(11):1707–1712.

3. Duffee DF, Shen W-K, Smith HC. Suppression of frequent premature ventricular contractions and improvement of left ventricular function in patients with presumed idiopathic dilated cardiomyopathy. *Mayo Clin Proc*. 1998;73(5):430–433.

4. Yarlagadda RK, Iwai S, Stein KM, et al. Reversal of cardiomyopathy in patients with repetitive monomorphic ventricular ectopy originating from the right ventricular outflow tract. *Circulation*. 2005;112(8):1092–1097.

5. Cha Y-M, Lee GK, Klarich KW, Grogan M. Premature ventricular contraction-induced cardiomyopathy. *Circ Arrhythm Electrophysiol*. 2012;5(1):229–236.

6. Bogun F, Crawford T, Reich S, et al. Radiofrequency ablation of frequent, idiopathic premature ventricular complexes: comparison with a control group without intervention. *Heart Rhythm*. 2007;4(7):863–867.

7. Baman TS, Lange DC, Ilg KJ, et al. Relationship between burden of premature ventricular complexes and left ventricular function. *Heart Rhythm*. 2010;7(7):865–869.

8. Taieb JM, Maury P, Shah D, et al. Reversal of dilated cardiomyopathy by the elimination of frequent left or right premature ventricular contractions. *J Interv Cardiac Electrophysiol*. 2007;20(1):9–13.

9. Kanei Y, Friedman M, Ogawa N, Hanon S, Lam P, Schweitzer P. Frequent premature ventricular complexes originating from the right ventricular outflow tract are associated with left ventricular dysfunction. *Ann Noninvasive Electrocardiol*. 2008;13(1):81–85.

10. Zhong L, Lee YH, Huang XM, et al. Relative efficacy of catheter ablation vs antiarrhythmic drugs in treating premature ventricular contractions: a single-center retrospective study. *Heart Rhythm*. 2014;11(2):187–193.

11. Deyell MW, Park KM, Han Y, et al. Predictors of recovery of left ventricular dysfunction after ablation of frequent ventricular premature depolarizations. *Heart Rhythm*. 2012;9(9):1465–1472.

12. Al-Khatib SM, Stevenson WG, Ackerman MJ, et al. 2017 AHA/ACC/HRS guideline for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. *J Am Coll Cardiol*. 2018;72(14):e91–e220.

13. Somani S, Russak AJ, Richter F, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. *Europace*. 2021;23(8):1179–1191.

14. Lampert J, Pulaski M, Miller MA, et al. Right precordial U waves in severe aortic stenosis: "real-world" prevalence and diminution by TAVR. *J Am Coll Cardiol*. 2022;79(24):2467–2469.

15. Nakamura T, Nagata Y, Nitta G, et al. Prediction of premature ventricular complex origins using artificial intelligence-enabled algorithms. *Cardiovasc Digit Health J*. 2021;2(1):76–83.

16. Zhao W, Zhu R, Zhang J, et al. Machine learning for distinguishing right from left premature ventricular contraction origin using surface electrocardiogram features. *Heart Rhythm*. 2022;19(11):1781–1789.

17. Zheng J, Fu G, Abudayyeh I, et al. A high-precision machine learning algorithm to classify left and right outflow tract ventricular tachycardia. *Front Physiol*. 2021;12:641066.

18. Zhou B, Khosla A, Lapedriza À, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas: NV, IEEE, 2016:2921–2929.

19. Ban JE, Park HC, Park JS, et al. Electrocardiographic and electrophysiological characteristics of premature ventricular complexes associated with left ventricular dysfunction in patients without structural heart disease. *Europace*. 2013;15(5):735–741.

20. Del Carpio Munoz F, Syed FF, Noheria A, et al. Characteristics of premature ventricular complexes as correlates of reduced left ventricular systolic function: study of the burden, duration, coupling interval, morphology and site of origin of PVCs. *J Cardiovasc Electrophysiol*. 2011;22(7):791–798.

21. Moulton KP, Medcalf T, Lazzara R. Premature ventricular complex morphology. A marker for left ventricular structure and function. *Circulation*. 1990;81(4):1245–1251.

22. Sun Y, Blom NA, Yu Y, et al. The influence of premature ventricular contractions on left ventricular function in asymptomatic children without structural heart disease: an echocardiographic evaluation. *Int J Cardiovasc Imaging*. 2003;19(4):295–299.

23. Bradfield JS, Homsi M, Shivkumar K, Miller JM. Coupling interval variability differentiates ventricular ectopic complexes arising in the aortic sinus of valsalva and great cardiac vein from other sources: mechanistic and arrhythmic risk implications. *J Am Coll Cardiol*. 2014;63(20):2151–2158.

24. Alhede C, Higuchi S, Hadjis A, et al. Premature ventricular contractions are presaged by a mechanically abnormal sinus beat. *J Am Coll Cardiol EP*. 2022;8(8):943–953.

25. Smith ML, Hamdan MH, Wasmund SL, Kneip CF, Joglar JA, Page RL. High-frequency ventricular ectopy can increase sympathetic neural activity in humans. *Heart Rhythm*. 2010;7(4):497–503.

26. Wilber DJ. Ventricular ectopic beats: not so benign. *Heart*. 2009;95(15):1209–1210.

27. Huizar JF, Kaszala K, Potfay J, et al. Left ventricular systolic dysfunction induced by ventricular ectopy: a novel model for premature ventricular contraction-induced cardiomyopathy. *Circ Arrhythm Electrophysiol*. 2011;4(4):543–549.

28. Jayakumar P, Moore MG, Furlough KA, et al. Comparison of an artificial intelligence–enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes in adults with knee osteoarthritis: a randomized clinical trial. *JAMA Netw Open*. 2021;4(2):e2037107.

29. Austin CA, Mohottige D, Sudore RL, Smith AK, Hanson LC. Tools to promote shared decision making in serious illness: a systematic review. *JAMA Intern Med*. 2015;175(7):1213–1221.

30. Bang V, Ganatra S, Shah SP, et al. Management of patients with giant cell myocarditis: JACC review topic of the week. *J Am Coll Cardiol*. 2021;77(8):1122–1134.

31. Kociol RD, Cooper LT, Fang JC, et al. Recognition and initial management of fulminant myocarditis: a scientific statement from the American Heart Association. *Circulation*. 2020;141(6):e69–e92.

32. Templin C, Ghadri JR, Diekmann J, et al. Clinical features and outcomes of takotsubo (stress) cardiomyopathy. *N Engl J Med*. 2015;373(10):929–938.

33. Shinbane JS, Wood MA, Jensen DN, Ellenbogen KA, Fitzpatrick AP, Scheinman MM. Tachycardia-induced cardiomyopathy: a review of animal models and clinical studies. *J Am Coll Cardiol*. 1997;29(4):709–715.

34. Medina de Chazal H, Del Buono MG, Keyser-Marcus L, et al. Stress cardiomyopathy diagnosis and treatment: JACC state-of-the-art review. *J Am Coll Cardiol*. 2018;72(16):1955–1971.

35. Manthey J, Imtiaz S, Neufeld M, Rylett M, Rehm J. Quantifying the global contribution of alcohol consumption to cardiomyopathy. *Popul Health Metr*. 2017;15(1):20.

KEY WORDS **cardiomyopathy, machine learning, premature ventricular contraction**