

4-18-2022

## Electrocardiogram-Based Machine Learning Emulator Model for Predicting Novel Echocardiography-Derived Phenogroups for Cardiac Risk-Stratification: A Prospective Multicenter Cohort Study

Heenaben B. Patel  
Naveena Yanamala  
Brijesh Patel  
Sameer Raina  
Peter D. Farjo  
Srinidhi Sunkara  
Márton Tokodi  
Nobuyuki Kagiya  
Grace Casacalang-Verzosa  
Partho P. Sengupta

Follow this and additional works at: <https://aah.org/jpcrr>



Part of the [Cardiology Commons](#), [Cardiovascular Diseases Commons](#), [Cardiovascular System Commons](#), [Community Health and Preventive Medicine Commons](#), [Equipment and Supplies Commons](#), and the [Other Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons](#)

### Recommended Citation

Patel HB, Yanamala N, Patel B, Raina S, Farjo PD, Sunkara S, Tokodi M, Kagiya N, Casacalang-Verzosa G, Sengupta PP. Electrocardiogram-based machine learning emulator model for predicting novel echocardiography-derived phenogroups for cardiac risk-stratification: a prospective multicenter cohort study. J Patient Cent Res Rev. 2022;9:98-107. doi: [10.17294/2330-0698.1893](https://doi.org/10.17294/2330-0698.1893)

Published quarterly by Midwest-based health system Advocate Aurora Health and indexed in PubMed Central, the Journal of Patient-Centered Research and Reviews (JPCRR) is an open access, peer-reviewed medical journal focused on disseminating scholarly works devoted to improving patient-centered care practices, health outcomes, and the patient experience.

# Electrocardiogram-Based Machine Learning Emulator Model for Predicting Novel Echocardiography-Derived Phenogroups for Cardiac Risk-Stratification: A Prospective Multicenter Cohort Study

Heenaben B. Patel, MBBS,<sup>1</sup> Naveena Yanamala, PhD, MS,<sup>1,2</sup> Brijesh Patel, DO,<sup>1</sup> Sameer Raina, MD, MBA,<sup>1</sup> Peter D. Farjo, MD,<sup>1</sup> Srinidhi Sunkara, BTech,<sup>1</sup> Márton Tokodi, MD,<sup>1,3</sup> Nobuyuki Kagiya, MD, PhD,<sup>1,4,5</sup> Grace Casacalang-Verzosa, MD,<sup>1</sup> Partho P. Sengupta, MD, DM<sup>1</sup>

<sup>1</sup>Division of Cardiology, West Virginia University Heart and Vascular Institute, Morgantown, WV; <sup>2</sup>Institute for Software Research, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA; <sup>3</sup>Heart and Vascular Center, Semmelweis University, Budapest, Hungary; <sup>4</sup>Department of Cardiovascular Biology and Medicine, Juntendo University, Tokyo, Japan; <sup>5</sup>Department of Digital Health and Telemedicine Research and Development, Juntendo University, Tokyo, Japan

<b>Purpose</b>	Electrocardiography (ECG)-derived machine learning models can predict echocardiography (echo)-derived indices of systolic or diastolic function. However, systolic and diastolic dysfunction frequently coexists, which necessitates an integrated assessment for optimal risk-stratification. We explored an ECG-derived model that emulates an echo-derived model that combines multiple parameters for identifying patient phenogroups at risk for major adverse cardiac events (MACE).
<b>Methods</b>	In this substudy of a prospective, multicenter study, patients from 3 institutions (n=727) formed an internal cohort, and the fourth institution was reserved as an external test set (n=518). A previously validated patient similarity analysis model was used for labeling the patients as low-/high-risk phenogroups. These labels were utilized for training an ECG-derived deep neural network model to predict MACE risk per phenogroup. After 5-fold cross-validation training, the model was tested on the reserved external dataset.
<b>Results</b>	Our ECG-derived model showed robust classification of patients, with area under the receiver operating characteristic curve of 0.86 (95% CI: 0.79–0.91) and 0.84 (95% CI: 0.80–0.87), sensitivity of 80% and 76%, and specificity of 88% and 75% for the internal and external test sets, respectively. The ECG-derived model demonstrated an increased probability for MACE in high-risk vs low-risk patients (21% vs 3%; P<0.001), which was similar to the echo-trained model (21% vs 5%; P<0.001), suggesting comparable utility.
<b>Conclusions</b>	This novel ECG-derived machine learning model provides a cost-effective strategy for predicting patient subgroups in whom an integrated milieu of systolic and diastolic dysfunction is associated with a high risk of MACE. ( <i>J Patient Cent Res Rev.</i> 2022;9:98-107.)
<b>Keywords</b>	surface electrocardiography; echocardiography; diastolic dysfunction; machine learning; topological data analysis

Cardiovascular disease is the leading cause of morbidity and mortality globally and enacts an estimated health care cost of more than \$200 billion in the United States annually.<sup>1</sup> Effective, economical, and personalized prevention and risk-stratification strategies

are imperative to mitigate this burden.<sup>2</sup> With each patient visit, based on symptomatology and provider preference, further testing with questionable cost-effectiveness may be ordered.<sup>3</sup> For instance, more than 30% of echocardiograms are performed outside of appropriate use criteria.<sup>4,5</sup>

To address these concerns, machine learning, a subfield of applied artificial intelligence, has been applied to extract features from simpler cost-effective tests like signal-processed surface electrocardiography (spECG).<sup>6-9</sup> These models have the potential to optimize downstream

Corresponding author: Partho P. Sengupta, Heart and Vascular Institute, 1 Medical Center Drive, Morgantown, WV 26506 (ps1165@rwjms.rutgers.edu)

testing by predicting individual echocardiographic (echo) parameters, computed tomography-derived coronary artery calcium scoring, or laboratory test features like hyperkalemia.<sup>6-9</sup> Recent advances in machine learning offer an emerging role for its use in integrating spECG and echo, with the potential for disease diagnosis, risk-stratification, and prediction of clinical outcomes.<sup>10</sup> Clinical risk-stratification strategies often require the integration of various clinical parameters; for example, optimal echo evaluation relies on expert insights that can integrate multiple parameters, such as 2-dimensional measurements and Doppler-derived parameters of left ventricular structure, systolic function, and diastolic function, for a given individual. To this end, we have recently described the role of unsupervised machine learning approaches using patient similarity analysis for integrating multiple echo parameters and phenogrouping patients with similar cardiac structure, systolic and diastolic function, and associated valvular heart diseases.<sup>11,12</sup> Using multiple external validation steps, these phenogroups were shown to be superior to conventional guidelines-based strategies for integrating echo parameters for predicting the risk of major adverse cardiac events (MACE).<sup>13,14</sup>

In the investigation presented herein, we sought to transfer the knowledge of echo-derived risk prediction to develop an electrocardiogram (ECG)-based prediction model that emulates echo-derived risk phenogroups of cardiac dysfunction.<sup>12</sup> We hypothesized that machine learning modeling using ECG and clinical parameters could effectively transcribe and approximate echo-based risk-stratification models for predicting patients at risk for MACE.

A parallel co-learning approach informed the development of a model using ECG, spECG, and clinical variables that emulates a model previously developed using echo variables.<sup>15</sup> Such multimodal co-learning approaches have been used in various techniques — emotion detection using electroencephalography and eye signals, audiovisual speech recognition devices for visually impaired persons (sensory signal and voice signal use), behavior, and action recognition for security reasons — to create a fusion model for better predictions.<sup>15,16</sup> A dataset where the information is originated from different sources can be considered multimodal, and the concept of parallel co-learning can be applied when data come from different modalities with overlapping instances,<sup>15</sup> ie, both ECG and echo data of the same patient. As a final validation step, we compared the performance of the ECG-based model with an independent, previously published echo-based model<sup>12</sup> to predict MACE.

## METHODS

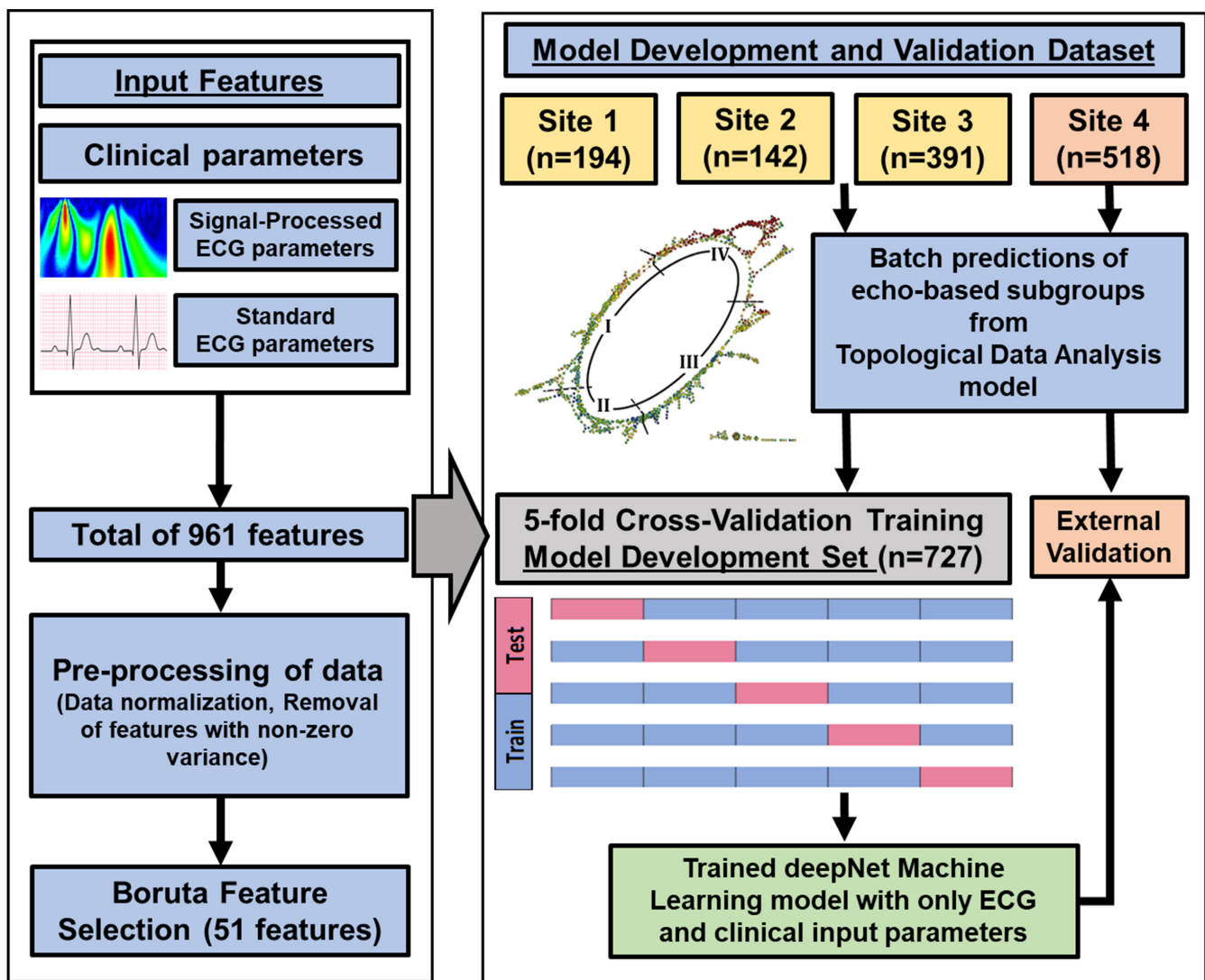
### Study Cohort

This study performed secondary analysis on data collected from a large trial. The patients included in this study were enrolled at the following institutions (n=1461): 1) Icahn School of Medicine at Mount Sinai in New York (NCT02560168); 2) the University of California, Los Angeles (UCLA) (NCT02873052); 3) Windsor Cardiac Center in Ontario, Canada; and 4) West Virginia University (WVU) in Morgantown, West Virginia.<sup>17</sup> The original trial and this study complied with the Declaration of Helsinki.<sup>18</sup> All sites received proper ethical oversight from WVU's institutional review board, and appropriate consent was obtained from study participants.

Subjects were screened before enrollment for the site-specific inclusion and exclusion criteria as detailed in Online Appendix A. This study aimed to develop a machine learning model using spECG, ECG, and clinical parameters to predict patient subgroups with high or low risk of MACE delineated using echo. The steps to develop this model for the study are outlined in Figure 1, and details of MyoVista® spECG (HeartSciences) can be found in Online Appendix A.

### Machine Learning Model Development

Data collected from three geographically separated institutions — Mount Sinai, UCLA, and Windsor Cardiac Center — were included in the final model development (n=727 patients). WVU served as an external validation dataset for further evaluation of the developed model (n=518 patients). As a first step, echo-derived high- or low-risk phenogroup labeling was performed using a topological data analysis (TDA) approach, described previously.<sup>12</sup> Briefly, 9 echo parameters, including ejection fraction (EF), left ventricular mass index, left atrial volume index (LAVi), early diastolic transmitral flow velocity (E), late diastolic transmitral flow velocity (A), E/A ratio, early diastolic relaxation velocity (e'), E/e' ratio, and tricuspid regurgitation peak velocity (TR V<sub>max</sub>) were used for TDA prediction of risk phenotypes. Patients were assigned low-risk (regions 1 and 2) or high-risk (regions 3 and 4) labels depending on their location in the TDA loop structure. These regions were found to have distinct left ventricular structure and functional echo parameters, along with a significant difference in MACE-related rehospitalization and death.<sup>12</sup> TDA was shown to identify data points associated with patients of similar phenotypic features from a multidimensional feature space and ultimately formed clusters on a similarity network. Enrollment characteristics of the prior study cohort used for the TDA model development are detailed in Online Appendix A. Thus, these previously validated echo-based risk phenotypes were used as a class label to train



**Figure 1.** Steps to develop a machine learning emulator model to predict echocardiographically defined patient subgroups. A total of 961 signal-processed electrocardiograms (ECGs), traditional ECGs, and clinical features were initially considered input features. Data were preprocessed and normalized. Boruta feature selection was performed using R, and only 51 significant features were used for model development. This was followed by obtaining echo-derived high-risk or low-risk labels for each patient through batch prediction from a previously validated model of topological data analysis (looped network of progressive cardiac dysfunction).<sup>12</sup> After training with 5-fold cross-validation, the model was subsequently tested on the reserved dataset for external validation.

and develop an ECG-based supervised machine learning model. Subjects with missing echo parameters to obtain TDA-derived risk group labels were excluded (n=216) from the analysis due to the nonfeasibility of training supervised machine learning for this group.

Clinical parameters such as age, sex, height, weight, body surface area, body mass index, heart rate, systolic blood pressure, diastolic blood pressure, and history of coronary artery disease, diabetes, hypertension, hyperlipidemia, and tobacco abuse, along with spECG (n=520) and ECG (n=427) parameters, were initially considered as input

features for the training of a supervised machine learning algorithm. We preprocessed our data to remove columns with zero variance, normalized all of the ECG data, and performed feature selection to select meaningful features using the Boruta algorithm in the R statistical environment to improve model accuracy.<sup>19,20</sup> Out of 961 clinical, spECG, and ECG features considered initially, the Boruta algorithm selected 51 significant features that were used for the machine learning model development. Boruta feature selection automatically identifies attributes from the dataset that are paramount for predicting outcomes and removing unnecessary features that have the least effect on the

same. This can be achieved by making shadow attributes, randomizing them, and calculating z-scores using a random forest equation, cross-referencing maximum z-scores to eliminate “unimportant” attributes, then repeating the process until all important features are identified.<sup>21</sup>

### ECG-Based Deep-Learning Classifier and Internal Validation

A deep neural network (DNN) is a collection of neurons where multiple layers are sequentially organized — starting as the input layer, followed by hidden layers, and, lastly, an output layer. Input layers bring the initial data for system processing in neurons. Each layer propagates information after adding weighted input, then transforms this information with nonlinear functions by mathematical computation.<sup>22</sup> As a result, it derives values as an output to the next layer for the subsequent neuron activation and makes an optimized prediction based on the precise neural network structure.

In this vein, ECG-based models were developed on the training set of clinical, spECG, and traditional ECG features employing a cloud-based automated machine learning platform (<http://bigml.com>, OptiML, BigML, Inc.). Numerous supervised machine learning algorithms, such as boosted trees, Deepnets (an optimized version of DNNs), random decision forest, bootstrap decision forest, and logistic regression models were trained and evaluated using 5-fold cross-validation to predict and classify patients who were labeled as high- or low-risk phenogroups using echo variables. To minimize the misprediction of high risk to the low-risk group due to class imbalance, the model development was optimized for the F-score of the high-risk group during the training process. Evaluation parameters such as accuracy, precision, recall, F-measure, and phi coefficient were used to compare various models and select the best-performing model.

In our evaluation analysis, auto network search, a type of Deepnets algorithm, was the best performing model generated by the BigML platform. This model assessed a total of 128 networks, and the auto network search optimization chose an ensemble of the top 7 best-performing networks for building the final predictive DNN model. The network search is not a random process but, in fact, a stepwise mathematical modeling method using Bayesian parameter optimization to identify hyperparameters during the search.

### External Validation Cohort and Follow-Up on Outcomes

To evaluate performance and overall generalizability of the developed DNN model, we used an unseen cohort of patients (test dataset) from the fourth institution (WVU)

not included during the model training process to determine the prediction accuracy. The patients in the external test cohort (n=518) were followed up for clinical outcomes, including cardiac death and MACE rehospitalization as defined by *International Classification of Diseases, Ninth/Tenth Revision* coding. The follow-up was censored at 38 months. Outcomes of interest for MACE include hospitalization for heart failure, myocardial ischemia, and revascularization (percutaneous coronary intervention or coronary artery bypass grafting); stroke; and cardiac death. The external validity of the DNN model for predicting clinical outcomes was evaluated.

### Statistical Analysis

We performed the Shapiro-Wilk test to check normality of the data. We used parametric methods for the variables that were normally distributed and nonparametric tests for nonnormal distributions for all statistical analyses. Continuous variables were expressed as mean  $\pm$  standard deviation or median (interquartile range), whereas categorical variables were presented as counts (percentages). Comparisons of continuous variables between the training and test set were performed using an independent sample *t*-test. If all cells of the contingency table contained 5 or more patients, then a chi-squared test was used to analyze categorical variables. If this assumption was not met, we performed a Fisher’s exact test. To evaluate whether both modalities led to similar clinical outcome predictions, we performed survival analysis on the follow-up data using Kaplan-Meier curves. In addition, a Cox proportional hazards regression model calculated a hazard ratio associated with MACE outcomes, time to the event, and predicted phenogroups. Medcalc for Windows 19.6 and R statistical analysis software was used for all statistical calculations. A P-value of <0.05 was considered statistically significant.

## RESULTS

### Study Population

The baseline clinical characteristics of the study cohort (training and test sets) are shown in Table 1. The patient population of the test dataset was significantly younger ( $P<0.0001$ ), had a greater body size ( $P<0.0001$ ), and had a significantly higher prevalence of coronary artery disease ( $P<0.0001$ ) than the training dataset. In contrast, patients in the training dataset had a higher prevalence of valvular heart disease including moderate-to-severe aortic, mitral, or tricuspid regurgitation ( $P<0.0001$ ), a higher prevalence of left ventricular hypertrophy ( $P<0.0001$ ), and more frequently had septal  $e'$  of  $<7$  cm/s or lateral  $e'$  of  $<10$  cm/s, TR  $V_{\max}$  of  $>2.8$  m/s, and LAVi of  $>34$  ml/m<sup>2</sup> ( $P<0.0001$  for all).



**Table 1.** Characteristics of the Internal (Training) and the External (Validation) Test Cohorts

Characteristic	Internal n=727	External n=518	P
Age, years	61 ± 14	50 ± 16	<0.0001*
Sex, male	406 (55.85%)	236 (45.56%)	0.0003*
Race			<0.0001*
White	406 (55.85%)	476 (91.89%)	
African-American	57 (7.84%)	5 (0.97%)	
Hispanic	35 (4.81%)	6 (1.16%)	
Asian	24 (3.30%)	17 (3.28%)	
Other or Unknown	205 (28.20%)	14 (2.70%)	
Body mass index, kg/m <sup>2</sup>	29.09 ± 6.08	32.25 ± 9.33	<0.0001*
Vitals			
Systolic blood pressure, mm Hg	130 ± 18	129 ± 18	0.46
Diastolic blood pressure, mm Hg	76 (69–83)	77 (70–84)	<0.0001*
Heart rate, beats/minute	66 ± 11	74 ± 14	<0.0001*
Comorbidities			
Diabetes mellitus	179 (24.62%)	108 (20.85%)	0.11
Hypertension	407 (55.98%)	316 (61.00%)	0.07
Hyperlipidemia	443 (60.94%)	306 (59.07%)	0.5
History of CAD, PCI, or CABG	46 (6.33%)	137 (26.45%)	<0.0001*
Echocardiography			
Reduced ejection fraction (<50%)	52 (7.15%)	35 (6.76%)	0.78
Left ventricular hypertrophy (concentric and eccentric)	161 (22.15%)	52 (10.04%)	<0.0001*
Aortic stenosis (moderate to severe)	19 (2.61%)	10 (1.93%)	0.43
Aortic regurgitation (moderate to severe)	44 (6.05%)	3 (0.58%)	<0.0001*
Mitral regurgitation (moderate to severe)	50 (6.88%)	7 (1.35%)	<0.0001*
Tricuspid regurgitation (moderate to severe)	40 (5.50%)	5 (0.97%)	<0.0001*
LVDD index			
Average E/e' > 14	60 (8.25%)	45 (8.69%)	0.78
Septal e' < 7 cm/s or lateral e' < 10 cm/s	455 (62.59%)	252 (48.65%)	<0.0001*
TR V <sub>max</sub> > 2.8 m/s	81 (11.14%)	13 (2.51%)	<0.0001*
LAVi > 34 ml/m <sup>2</sup>	166 (22.83%)	69 (13.32%)	<0.0001*
LVDD (2 or more criteria met)	151 (20.77%)	76 (14.67%)	0.006*

Values are presented as counts (%), mean ± standard deviation, or, in the case of nonnormal distribution for diastolic blood pressure, median (interquartile range). P-values were calculated using an independent t-test where the mean is reported and chi-squared or Fisher's exact test where frequencies are reported.

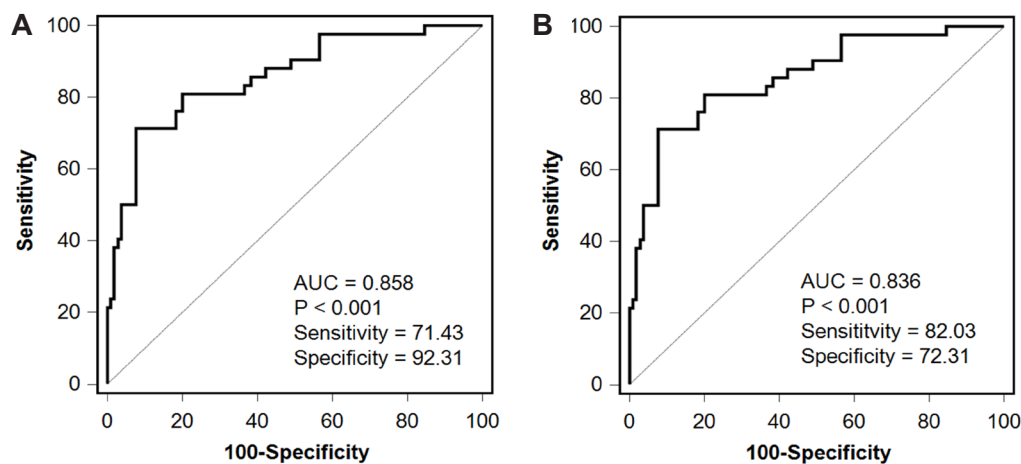
\*P<0.05 indicates significant difference between training and external test set.

CAD, coronary artery disease; PCI, percutaneous coronary intervention; CABG, coronary artery bypass graft; E, early diastolic mitral wave velocity; e', tissue Doppler-derived early diastolic mitral annular velocity; LAVi, left atrial volume index; LVDD, left ventricular diastolic dysfunction; TR V<sub>max</sub>, tricuspid regurgitation peak velocity.

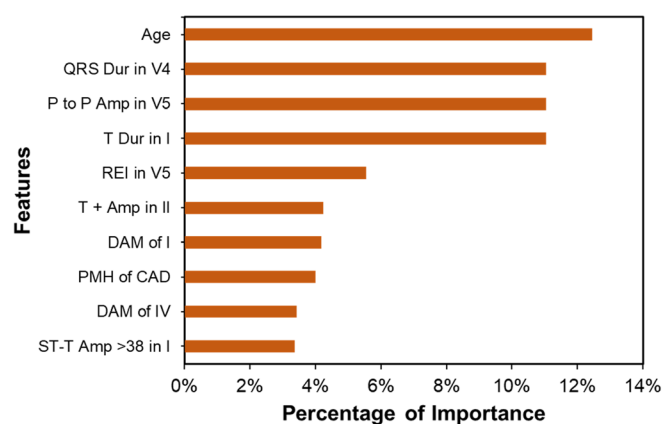
### Performance Evaluation of the ECG-Based DNN

Despite the inherent difference in the population characteristics and geographical location, the developed predictive model was able to effectively identify and predict TDA-derived echo-based patient subgroups at low or high risk of MACE. The DNN model showed robust classification of patients with areas under the receiver operating characteristic curves (AUC) of 0.86

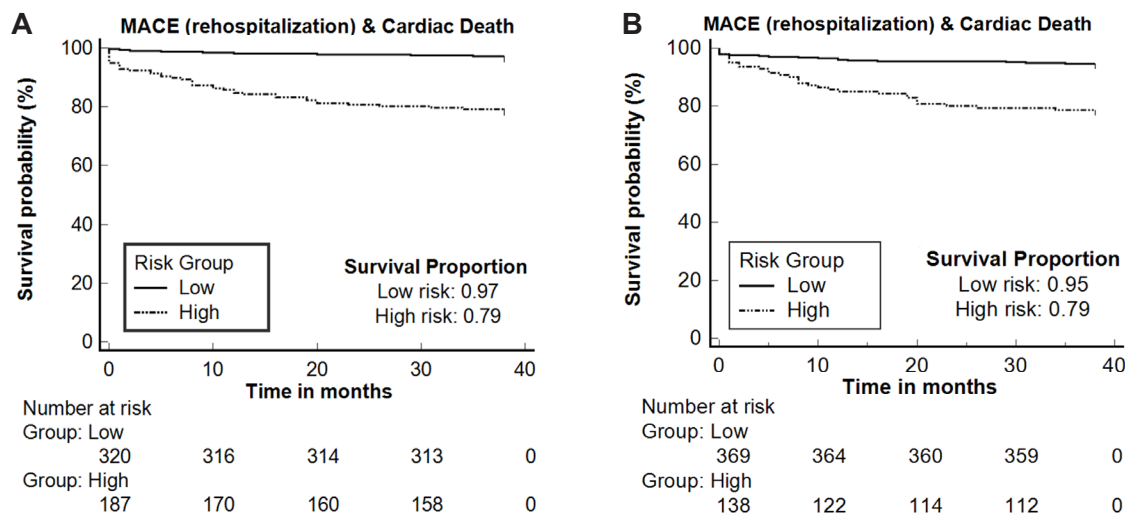
(95% CI: 0.79–0.91) and 0.84 (95% CI: 0.80–0.87), sensitivities of 80% and 76%, and specificities of 88% and 75% for the internal and external test datasets, respectively (Figure 2). Here, an AUC number of 0.86 indicates that there is an 86% chance that the model will be able to distinguish between positive class (high risk for MACE events) and negative class (low risk for MACE events) for the internal test dataset and, similarly,



**Figure 2.** Receiver operating characteristic curves for the deep neural network model predicting echocardiographically defined patient subgroups at low or high risk of major adverse cardiac events for 5-fold cross-validation (A) and external validation (B). An area under the curve (AUC) of >0.5 indicates better predictive values.



**Figure 3.** The importance of features within the developed deep neural network model. Of the 51 features selected, 51% came from signal-processed electrocardiograms, 45% from traditional electrocardiograms, and 4% from clinical features. The details of these features are described in Online Supplemental Table S1. Amp, amplitude; PMH of CAD, past medical history of coronary artery disease; DAM, depolarization average measure; Dur, duration; REI, repolarization early minimum.



**Figure 4.** Kaplan-Meier curves for major adverse cardiovascular events (MACE) in the test set and for time-to-event analysis of the external test set. The outcomes of interest include events such as MACE-related rehospitalization and cardiac death in the external validation cohort that occurred throughout the 3-year follow-up period. All patients were censored at study follow-up at 38 months. Survival analysis was done using predicted probabilities for high-risk and low-risk groups for signal-processed electrocardiographic (spECG), traditional electrocardiographic, and clinical feature-based machine learning model (A), and for directly echocardiographic (echo) parameter-trained machine learning (B);  $P < 0.0001$  on log-rank test for both plots. The plots demonstrate similar performance for the high-risk group (survival proportion of 0.79 for A and B) and slightly improved performance for a low-risk group [survival proportion of 0.97 for A and 0.95 for B) using the spECG model when compared with the directly echo-trained model.

84% chance for the external dataset; the higher the value of AUC, the better the model is at predicting negative classes as negative and positive classes as positive. The slightly lower performance of the model in the external validation group could be attributable to the underlying disease status of its heterogeneous population from a distinct geographic location.

Evaluation plots corresponding to the external validity of the DNN model for predicting clinical outcomes are shown in Figure 2. The performance metrics of the model are shown in Table 2. Moreover, we have compared the performance of this model to a model trained only using standard ECG data, and the performance findings were AUC of 0.79 for internal and 0.75 for the external test set using 80% data for training and 20% data as a testing dataset. The model was tested extensively, and the very low performance found using only standard ECG features suggests that the addition of signal-processed ECG data is informative and helpful for predicting outcomes of interest.

### Features of Importance

Of 961 available features (14 clinical, 520 spECG, and 427 ECG), 51 features selected by the Boruta algorithm came mainly from spECG (n=26, 51%), followed by traditional ECG (n=23, 45%) and clinical features (n=2, 4%). The list of the 51 Boruta-selected features is summarized in Online Supplemental Table S1. Among the 51 features, the most important for classification were the clinical features age and past medical history of coronary artery disease, the spECG features depolarization average measure in lead I and IV and repolarization early minimum in lead V5, and the conventional ECG features QRS duration in lead V4, P-to-P amplitude in V5, T duration in lead I, T amplitude in lead II (positive) and ST-T amplitude of >38 mv in lead I — collectively, these measures provided a robust prediction of MACE in the external test set.

The top 10 features are shown in Figure 3. Interestingly, age (12.45%) was the most important among all selected features.

### Time-to-Event Survival Analysis

The ECG-predicted model demonstrated an increased probability for MACE in high-risk compared to low-risk patients (21% vs 3%;  $P<0.001$ ); the results were similar to the echo-trained model (21% vs 5%;  $P<0.001$ ), suggesting a comparable utility of ECG instead of echo to identify patients at risk of MACE events (Figure 4). In addition, the patients labeled as high risk using the ECG-trained model had less chance of survival than the low-risk group (Figure 4A; log-rank  $P<0.0001$ ). These results were similar or almost identical to the echo-trained model (Figure 4B; log-rank  $P<0.0001$ ).

**Table 2.** Performance of Deep Neural Network Prediction Model on Training and External Datasets

Metric	5-fold cross-validation	External validation
AUC ROC <sup>a</sup>	0.86 (95% CI: 0.79–0.91)	0.84 (95% CI: 0.80–0.87)
Accuracy <sup>b</sup>	0.84	0.75
F-measure <sup>c</sup>	0.80	0.71
Precision <sup>d</sup>	0.80	0.70
Recall <sup>e</sup>	0.80	0.76

<sup>a</sup>Area under the receiver operating characteristic curve.

<sup>b</sup>Proportion of correct predictions out of total predictions that were generated.

<sup>c</sup>F-score is a weighted mean considering both precision and recall measures.

<sup>d</sup>Proportion of true positive predicted values identified out of all true-positive and false-positive values.

<sup>e</sup>Proportion of actual positive values identified from total true-positive and false-negative values.

In the Cox proportional-hazards regression analyses, an increased risk of MACE was observed for high-risk phenogroups (hazard ratio: 8.17 [95% CI: 3.97 to 16.82];  $P<0.0001$ ) when compared with patients that belong to low-risk phenogroups.

## DISCUSSION

Recent deep-learning approaches have utilized large retrospective ECG datasets for developing generalizable machine learning models for predicting left ventricular systolic function.<sup>23</sup> In contrast, we used a novel emulator-model approach wherein, despite a smaller training sample size, we were able to achieve a high level of performance for identifying echo-based risk phenotypes. These risk phenotypes were shown to successfully integrate left ventricular structure and function for predicting future MACE.<sup>12</sup> Both the training and testing were done using a prospectively obtained database of more than 1200 participants from multiple centers located in North America. Notably, generalizability of the model also was tested using an external test site, as ability to adapt properly to new data is central to the success of any predictive machine learning model.<sup>24</sup> Moreover, we trained the ECG model for identifying echo-based risk phenotype. Finally, in addition to predicting phenogroups, we tested the model performance for predicting future MACE with 3 years of clinical follow-up data. These findings confirm our hypothesis that knowledge developed on an echo-derived model can be successfully transferred to an ECG-derived model by using a co-learning technique. Furthermore, the survival analysis comparing ECG-trained vs echo-trained



models demonstrated almost identical results for MACE in high-risk patients, suggesting the potential use of ECG modality as a valuable alternative cost-effective risk-stratification tool.

Several studies have demonstrated the utility of standard ECG parameters for predicting clinical cardiac outcomes with or without using machine learning algorithms.<sup>25-31</sup> A prospective study conducted by Al-Zaiti et al in three tertiary U.S. hospitals demonstrated the utility of machine learning-based prediction of acute coronary syndrome using 12-lead standard ECG.<sup>25</sup> Similarly, myocardial infarction, stroke, and mortality predictors were identified using ECG.<sup>26,27</sup> While the use of single parameters like ejection fraction can predict the development of heart failure, considerable bias may be introduced in diagnosing overall categories of heart failure syndromes that present with varying extents of systolic and diastolic dysfunction. However, chronic heart failure is a complex, multifactorial syndrome consisting of many overlapping phenotypes,<sup>32</sup> in which the features of structural remodeling and systolic and diastolic dysfunction slowly progress from subclinical stages toward the development of overt heart failure.

Since heart failure does not emerge as a uniform phenotype, but instead as a disease spectrum of overlapping phenotypes, we previously investigated this continuum using patient similarity analysis. Specifically, we used TDA, a fundamental advancement in machine learning, which demonstrated the importance of understanding the “shape” of data to extract meaningful insights.<sup>33</sup> This technology allowed precise phenotypic recognition of the continuum of left ventricular response patterns during the progression of heart failure.<sup>11</sup> In the present study, we used the previous knowledge of the echo-derived TDA model to train our surrogate ECG-derived model. Several ECG features, including QRS duration, P-to-P amplitude, T duration/amplitude, and ST-T amplitude were among the most important variables for discriminating between high-risk and low-risk phenogroups. Intraventricular conduction delay and widening of QRS complex, QRT/T angle, QT-T durations, and ST-depressions are among several markers that predicted cardiovascular events in large-population studies.<sup>34</sup>

The addition of spECG parameters in the present study was done to extract and identify robust meaningful information from the electrical signal activities that lie in a large amount of wavelet-transformed cardiac energy data captured using MyoVista spECG. The quantitative values of cardiac energy at various time points of the cardiac cycle, along with frequency and amplitude data for a specific wave, generated more than 500 features to feed machine learning algorithms. The eventual machine

learning algorithm implemented was able to detect meaningful information from the ECG as well as spECG signals linked to the echo-derived TDA phenogroups.

## Limitations

This study had a relatively short-term follow-up to predict MACE. Perhaps a similar survey with longer follow-up could provide a cross-over of the risk groups and further shed light on the change in prognostication over time. Our analysis did not account for differences in spECG over time with treatment or lack thereof. The assessment of dynamic changes in spECGs and echocardiograms with time could provide essential clues about the interrelationship of structural and electrophysiologic changes in the heart that ultimately lead to MACE. Additionally, we did not use ECG nor spECG variables to predict MACE directly. Although the latter strategy could be employed independent of echo variables, it would endanger the potential value of explanation offered when associating with echo features of cardiac structure and function. Nevertheless, future strategies to combine both ECG and echo variables deserve important considerations in larger patient databases.

## CONCLUSIONS

Utilizing a wide spectrum of data — traditional and signal-processed ECG, patient demographics, and comorbidities — successfully predicted echocardiographically defined patient subgroups at high risk of major adverse cardiovascular events. Results demonstrate the potential value of machine learning-driven algorithms for rapid decision-making in an office-based setting to evaluate and monitor the progress of the patient and justify appropriate downstream referral for additional tests like echocardiography or other interventions.

### Patient-Friendly Recap

- An integrated assessment of heart function using clinical and electrocardiographic (ECG) features found to be reflective of echocardiographic analysis could help group patients by their risk for major cardiac events.
- Authors developed an ECG-based machine learning model to mirror echo-derived measures of function.
- They found that a spectrum of ECG features, patient demographics, and comorbidities was predictive of high or low risk as defined by echo.
- Refined ECG-based risk-stratification can provide a cost-effective strategy for classifying patients by their underlying and clinically meaningful cardiac dysfunction.

## Author Contributions

Study design: Yanamala, Sengupta. Data acquisition or analysis: H.B. Patel, Yanamala, Sunkara, Sengupta. Manuscript drafting: H.B. Patel, Yanamala, B. Patel, Raina, Farjo, Tokodi, Sengupta. Critical revision: H.B. Patel, Yanamala, B. Patel, Raina, Farjo, Tokodi, Kagiyaama, Casaclang-Verzosa, Sengupta.

## Conflicts of Interest

Partho Sengupta is a consultant for Kencor Health, Ultromics, and RCE technologies. All other authors have no conflicts of interest to disclose.

## Funding Sources

This work is supported in part by funds from the National Science Foundation (NSF: #1920920) and by Heart Test Laboratories, Inc. d/b/a HeartSciences. HeartSciences provided funding and signal-processed electrocardiography (spECG) devices. It had no role in developing the research plan, analysis, or manuscript other than providing necessary resources to collect the information from different site investigators.

## References

1. Weir HK, Anderson RN, Coleman King SM, et al. Heart disease and cancer deaths – trends and projections in the United States, 1969-2020. *Prev Chronic Dis*. 2016;13:E157. [CrossRef](#)
2. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. *Circulation*. 2019;140:e596-646. [CrossRef](#)
3. Karády J, Mayrhofer T, Ivanov A, et al. Cost-effectiveness analysis of anatomic vs functional index testing in patients with low-risk stable chest pain. *JAMA Netw Open*. 2020;3(12):e2028312. [CrossRef](#)
4. Bhatia RS, Ivers N, Yin CX, et al. Design and methods of the Echo WISELY (Will Inappropriate Scenarios for Echocardiography Lessen Significantly) study: an investigator-blinded randomized controlled trial of education and feedback intervention to reduce inappropriate echocardiograms. *Am Heart J*. 2015;170:202-9. [CrossRef](#)
5. Kirkpatrick JN, Ky B, Rahmouni HW, et al. Application of appropriateness criteria in outpatient transthoracic echocardiography. *J Am Soc Echocardiogr*. 2009;22:53-9. [CrossRef](#)
6. Farjo PD, Yanamala N, Kagiyaama N, et al. Prediction of coronary artery calcium scoring from surface electrocardiogram in atherosclerotic cardiovascular disease: a pilot study. *Eur Heart J Digit Health*. 2020;1:51-61. [CrossRef](#)
7. Sengupta PP, Kulkarni H, Narula J. Prediction of abnormal myocardial relaxation from signal processed surface ECG. *J Am Coll Cardiol*. 2018;71:1650-60. [CrossRef](#)
8. Attia ZI, DeSimone CV, Dillon JJ, et al. Novel bloodless potassium determination using a signal-processed single-lead ECG. *J Am Heart Assoc*. 2016;5(1):e002746. [CrossRef](#)
9. Yasin OZ, Attia Z, Dillon JJ, et al. Noninvasive blood potassium measurement using signal-processed, single-lead ECG acquired from a handheld smartphone. *J Electrocardiol*. 2017;50:620-5. [CrossRef](#)
10. Chang A, Cadaret LM, Liu K. Machine learning in electrocardiography and echocardiography: technological advances in clinical cardiology. *Curr Cardiol Rep*. 2020;22(12):161. [CrossRef](#)
11. Casaclang-Verzosa G, Shrestha S, Khalil MJ, et al. Network tomography for understanding phenotypic presentations in aortic stenosis. *JACC Cardiovasc Imaging*. 2019;12:236-48. [CrossRef](#)
12. Tokodi M, Shrestha S, Bianco C, et al. Interpatient similarities in cardiac function: a platform for personalized cardiovascular medicine. *JACC Cardiovasc Imaging*. 2020;13:1119-32. [CrossRef](#)
13. Pandey A, Kagiyaama N, Yanamala N, et al. Deep-learning models for the echocardiographic assessment of diastolic dysfunction. *JACC Cardiovasc Imaging*. 2021;14:1887-900. [CrossRef](#)
14. Sengupta PP, Shrestha S, Kagiyaama N, et al. A machine-learning framework to identify distinct phenotypes of aortic stenosis severity. *JACC Cardiovasc Imaging*. 2021;14:1707-20. [CrossRef](#)
15. Baltrušaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell*. 2019;41:423-43. [CrossRef](#)
16. Zheng WL, Dong BN, Lu BL. Multimodal emotion recognition using EEG and eye tracking data. *Annu Int Conf IEEE Eng Med Biol Soc*. 2014;2014:5040-3. [CrossRef](#)
17. Kagiyaama N, Piccirilli M, Yanamala N, et al. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. *J Am Coll Cardiol*. 2020;76:930-41. [CrossRef](#)
18. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310:2191-4. [CrossRef](#)
19. Breiman L. Random forests. *Machine Learning*. 2001;45:5-32. [CrossRef](#)
20. Kursu MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw*. 2010;36(11):1-13. [CrossRef](#)
21. Manhar MA, Soesanti I, Setiawan NA. Improving feature selection on heart disease dataset with Boruta approach. *Journal FORTEI-JEERI*. 2020;1(1):41-48.
22. Montavon G, Samek W, Müller KR. Methods for interpreting and understanding deep neural networks. *Digit Signal Process*. 2018;73:1-15. [CrossRef](#)
23. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25:70-74. [CrossRef](#)
24. Sengupta PP, Shrestha S, Berthon B, et al. Proposed requirements for cardiovascular imaging-related machine learning evaluation (PRIME): a checklist. *JACC Cardiovasc Imaging*. 2020;13:2017-35. [CrossRef](#)
25. Al-Zaiti S, Besomi L, Bouzid Z, et al. Machine learning-based prediction of acute coronary syndrome using only the pre-hospital 12-lead electrocardiogram. *Nat Comm*. 2020;11(1):3966. [CrossRef](#)
26. Rautaharju PM, Kooperberg C, Larson JC, LaCroix A. Electrocardiographic abnormalities that predict coronary heart disease events and mortality in postmenopausal women: the Women's Health Initiative. *Circulation*. 2006;113:473-80. [CrossRef](#)

27. Christensen H, Fogh Christensen A, Boysen G. Abnormalities on ECG and telemetry predict stroke outcome at 3 months. *J Neurol Sci.* 2005;234:99-103. [CrossRef](#)
28. Deo R, Shou H, Soliman EZ, et al. Electrocardiographic measures and prediction of cardiovascular and noncardiovascular death in CKD. *J Am Soc Nephrol.* 2016;27:559-69. [CrossRef](#)
29. Hedén B, Öhlin H, Rittner R, Edenbrandt L. Acute myocardial infarction detected in the 12-lead ECG by artificial neural networks. *Circulation.* 1997;96:1798-802. [CrossRef](#)
30. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf Sci.* 2017;415-416:190-8. [CrossRef](#)
31. Forberg JL, Green M, Björk J, et al. In search of the best method to predict acute coronary syndrome using only the electrocardiogram from the emergency department. *J Electrocardiol.* 2009;42(1):58-63. [CrossRef](#)
32. De Keulenaer GW, Brutsaert DL. Systolic and diastolic heart failure are overlapping phenotypes within the heart failure spectrum. *Circulation.* 2011;123:1996-2004. [CrossRef](#)
33. Lum PY, Singh G, Lehman A, et al. Extracting insights from the shape of complex data using topology. *Sci Rep.* 2013;3:1236. [CrossRef](#)
34. Zhang ZM, Prineas RJ, Case D, Soliman EZ, Rautaharju PM; ARIC Research Group. Comparison of the prognostic significance of the electrocardiographic QRS/T angles in predicting incident coronary heart disease and total mortality (from the Atherosclerosis Risk in Communities study). *Am J Cardiol.* 2007;100:844-9. [CrossRef](#)

© 2022 Advocate Aurora Health, Inc.