Circulation: Arrhythmia and Electrophysiology

ORIGINAL ARTICLE

Machine Learning Prediction of Response to Cardiac Resynchronization Therapy

Improvement Versus Current Guidelines

BACKGROUND: Cardiac resynchronization therapy (CRT) has significant nonresponse rates. We assessed whether machine learning (ML) could predict CRT response beyond current guidelines.

METHODS: We analyzed CRT patients from Cleveland Clinic and Johns Hopkins. A training cohort was created from all Johns Hopkins patients and an equal number of randomly sampled Cleveland Clinic patients. All remaining patients comprised the testing cohort. Response was defined as ≥10% increase in left ventricular ejection fraction. ML models were developed to predict CRT response using different combinations of classification algorithms and clinical variable sets on the training cohort. The model with the highest area under the curve was evaluated on the testing cohort. Probability of response was used to predict survival free from a composite end point of death, heart transplant, or placement of left ventricular assist device. Predictions were compared with current guidelines.

RESULTS: Nine hundred twenty-five patients were included. On the training cohort (n=470: 235, Johns Hopkins; 235, Cleveland Clinic), the best ML model was a naive Bayes classifier including 9 variables (QRS morphology, QRS duration, New York Heart Association classification, left ventricular ejection fraction and end-diastolic diameter, sex, ischemic cardiomyopathy, atrial fibrillation, and epicardial left ventricular lead). On the testing cohort (n=455, Cleveland Clinic), ML demonstrated better response prediction than guidelines (area under the curve, 0.70 versus 0.65; P=0.012) and greater discrimination of event-free survival (concordance index, 0.61 versus 0.56; P<0.001). The fourth quartile of the ML model had the greatest risk of reaching the composite end point, whereas the first quartile had the least (hazard ratio, 0.34; P<0.001).

CONCLUSIONS: ML with 9 variables incrementally improved prediction of echocardiographic CRT response and survival beyond guidelines. Performance was not improved by incorporating more variables. The model offers potential for improved shared decision-making in CRT (online calculator: http://riskcalc.org:3838/CRTResponseScore). Significant remaining limitations confirm the need to identify better variables to predict CRT response.

VISUAL OVERVIEW: A visual overview is available for this article.

Albert K. Feeny, BS
John Rickard, MD, MPH*
Divyang Patel, MD
Saleem Toro, MD
Kevin M. Trulock, MD
Carolyn J. Park, MD
Michael A. LaBarbera, MS
Niraj Varma, MD, PhD
Mark J. Niebauer, MD, PhD
Sunil Sinha, MD
Eiran Z. Gorodeski, MD, MPH
Richard A. Grimm, DO
Xinge Ji, MS
John Barnard, PhD
Anant Madabhushi, PhD*
David D. Spragg, MD*
Mina K. Chung, MD*

*Drs Rickard, Madabhushi, Spragg, and Chung are joint senior authors.

Key Words: algorithms ■ cardiac resynchronization therapy ■ heart failure ■ machine learning ■ patient selection

https://www.ahajournals.org/journal/circep

WHAT IS KNOWN?

- Cardiac resynchronization therapy improves heart failure outcomes but has significant nonresponse rates.
- Combined assessment of clinical variables with machine learning has the potential to improve cardiac resynchronization therapy response prediction.

WHAT THE STUDY ADDS?

- A machine learning model using 9 common clinical variables predicted reverse remodeling and long-term survival after cardiac resynchronization therapy with improvement versus current guidelines.
- Adding comorbidities and pharmacotherapy variables to the machine learning model did not improve its performance.
- The machine learning model could be used as a shared decision-making tool in cardiac resynchronization therapy with an online calculator available at http://riskcalc.org:3838/CRTResponseScore/.

Cardiac resynchronization therapy (CRT) leads to beneficial left ventricular (LV) remodeling and improved outcomes in patients with heart failure.[1–3] However, 20% to 50% of CRT patients have an apparent lack of positive response to therapy,[4] indicating the need for improved patient selection. The strongest indications for CRT include QRS duration (QRSd) ≥150 ms and the presence of left bundle branch block (LBBB).[5] Studies have identified other clinical variables that are associated with positive CRT response, such as female sex,[6] nonischemic cardiomyopathy,[7,8] and reduced comorbidity burden.[9]

Combined assessment of variables may improve CRT response prediction. Risk stratification via assessment of clinical variables improved prognosis prediction with CRT implantation[7] and within patients who had already received CRT devices.[10] As large electronic datasets have become available, machine learning (ML) has emerged as an exciting avenue for medical prediction problems,[11] with potential to assist CRT patient selection and shared decision-making. Cikes et al[12] used unsupervised ML of complex echocardiogram data and clinical parameters to identify heart failure phenogroups with differential CRT response. Kalscheur et al[13] provided insight into the potential for supervised ML models to predict all-cause mortality or heart failure hospitalization in a randomized controlled CRT trial, using a broad set of 45 common clinical variables.

The utility of supervised ML in CRT has important unanswered questions. Which outcomes can ML predict? What data are required to achieve these predictions? In this study, we sought to answer these questions and build on the findings of Kalscheur et al.[13] In observational patient cohorts, we evaluated the capability of ML analysis of clinical variables to improve CRT response prediction, defined by reverse cardiac remodeling measured on echocardiogram. We also assessed whether the ML prediction of reverse remodeling predicted mortality. We attempted to discern which data are required for optimal CRT response prediction by examining the effect of incorporating the following 7 different feature sets into the ML prediction models: (1) a minimal feature set of 9 handpicked variables (QRS morphology, QRSd, New York Heart Association [NYHA] classification, LV ejection fraction [LVEF], and LV end-diastolic diameter, sex, ischemic cardiomyopathy, atrial fibrillation, and epicardial LV lead), (2) a feature set that added comorbidity data, (3) a broad spectrum feature set that added pharmacotherapy, and (4–7) feature sets that were algorithmically derived using 4 different feature selection algorithms. We compared ML model predictions with guideline criteria and developed an online prediction calculator that can be used as a shared decision-making tool.

METHODS

The study was approved by the Institutional Review Boards of the Cleveland Clinic and Johns Hopkins for retrospective medical records review. Because of the sensitive nature of the data collected for this study, requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to the co-senior authors at Cleveland Clinic and Johns Hopkins. Additional institutional review board approval for data release would need to be obtained.

Study Population

We reviewed patients receiving CRT implants from 2003 to 2012 at the Cleveland Clinic and from 2002 to 2010 at Johns Hopkins Hospital. Patients with an echocardiographic LVEF measurement before CRT implant and an LVEF measurement at least 60 days after CRT implant were included for review. Patients missing clinical data required for model development were excluded. Echocardiograms were performed as clinically indicated, and the echocardiogram closest to 1-year follow-up was selected as the post-CRT echocardiogram. Positive CRT response was defined by at least 10% absolute increase in LVEF from baseline. We chose a high threshold for response to reduce identification of false relationships between variables and CRT response.

Study Design Overview

We used supervised ML classification to predict CRT response (Figure I in the Data Supplement). We followed a template for unbiased development of ML classifiers (Figure 1). The data were split into training and testing data. Training data were used to evaluate ML classifiers created from different combinations of feature sets and classifier algorithms in predicting echocardiographic CRT response. The highest

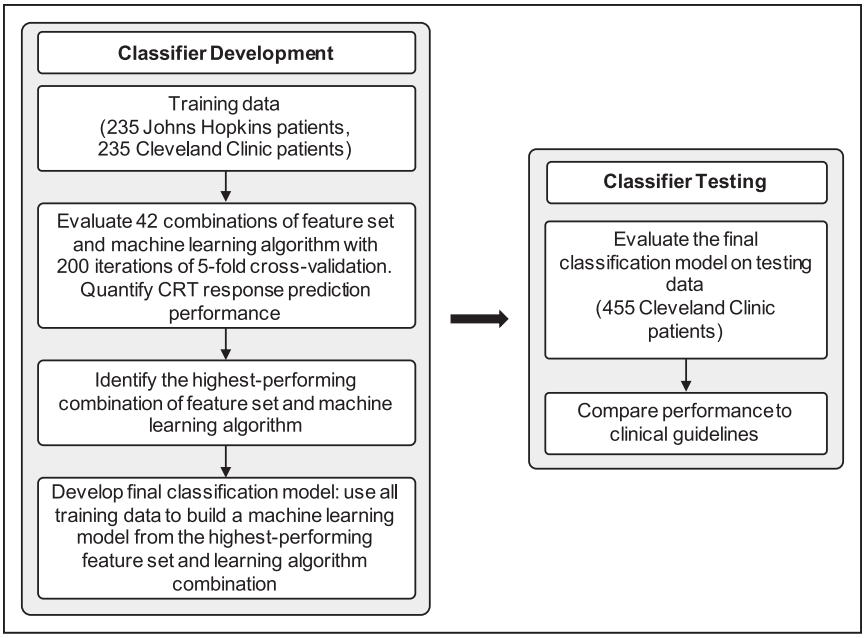


Figure 1. Machine learning classifier development and testing.
CRT indicates cardiac resynchronization therapy.

performing combination of feature set and learning algorithm was used to develop a final model from all training data, which was then evaluated on the testing data.

Data Preparation

Training data consisted of all patients from the Johns Hopkins cohort and an equal quantity of randomly sampled patients from the Cleveland Clinic cohort, such that the training set had equal representation from both institutions to limit institutional bias. Testing data were comprised by remaining patients from the Cleveland Clinic cohort. Noncategorical data were normalized by mean and SD.

Classifier Development

The goal of classifier development was to use cross-validation (Figure II in the Data Supplement) on the training cohort to identify the optimal feature set and learning algorithm combination to predict CRT response. We evaluated several classification algorithms (logistic regression, diagonal linear discriminant analysis, support vector machine, naive Bayes, random forest, and AdaBoost), each evaluated in combination with 7 different feature sets. The feature set–learning algorithm combination with the highest area under the curve (AUC) of the receiver operating characteristic (ROC) curve was selected to develop a final classifier for evaluation on the testing data. An operating point was specified by maximizing the Youden index on the mean ROC curve.[14] Further details are provided in Methods in the Data Supplement.

Feature Selection

The goal of feature selection was to identify the optimal subset of features useful in predicting CRT response. Feature selection informs which features are informative in prediction and can enhance prediction performance by removing nonpredictive features. We evaluated 7 different feature sets composed of commonly available clinical variables.

The first 3 feature sets were human derived. The first set, termed minimal, was a set of 9 features consisting of sex[6,15] and other cardiac characteristics identified as CRT response predictors[16]: QRSd,[7] QRS morphology (LBBB versus right bundle branch block versus nonspecific intraventricular conduction delay versus right ventricular pacing),[17] cause of cardiomyopathy (ischemic or nonischemic),[7] LVEF, LV end-diastolic diameter,[7] NYHA heart failure classification,[8] history of atrial fibrillation,[16] and use of surgical epicardial LV lead.[18] The second set, termed comorbidities, had 26 features, adding physical characteristics and comorbidities: age, height, weight, body mass index, hypertension, hyperlipidemia, chronic obstructive pulmonary disease, prior cerebrovascular accident or transient ischemic attack, malignancy, diabetes mellitus, serum creatinine, white blood cell count, hemoglobin, red blood cell distribution width, prior coronary artery bypass graft, prior percutaneous coronary intervention, and atrioventricular node ablation. The third set, termed broad, had 37 features and added pharmacological use: β-blocker, angiotensin-converting enzyme inhibitor or angiotensin receptor blocker, diuretic, nitrate, hydralazine, aldosterone antagonist, statin, warfarin, clopidogrel, digoxin, or antiarrhythmic therapy.

The fourth through seventh feature sets were algorithmically derived using 4 feature selection algorithms: minimum redundancy and maximum relevance,[19] random forest variable importance, univariate statistical testing (2-sample *t* test for continuous variables and χ^2 test for categorical variables[20]), and Lasso regularization.[9] Details are provided in Methods in the Data Supplement.

Classifier Testing

The goal of classifier testing was to evaluate the final ML classifier developed from the optimal feature set–learning algorithm combination identified in cross-validation. The ROC of the ML classifier was compared with the ROC constructed from 2013 American College of Cardiology Foundation/American Heart Association class I, IIa, and IIb guideline criteria[5] and the

ROC constructed using a simple modified CRT response score derived from sex, nonischemic cardiomyopathy, LBBB, QRSd ≥150 ms, and LV end-diastolic diameter[7] (Methods in the Data Supplement). Accuracy, sensitivity, specificity, positive predictive value, and negative predictive value at the operating point of the ML classifier were compared with guideline criteria, as well as to the predictive performance of nonischemic cardiomyopathy, QRSd ≥150 ms, LBBB, LBBB with QRSd ≥150 ms, and female sex.

We evaluated whether the ML Response Score, defined by the ML probability of response, was predictive of other echocardiographic outcomes. Two additional ROCs were constructed for a lower threshold for response (LVEF improvement, ≥5%) and for super response (LVEF improvement, ≥20%). To assist in ML interpretation, we computed descriptive statistics of patient subgroups defined by ML Response Score quartiles.

Survival Analysis

We sought to determine whether the ML Response Score predicted long-term survival free from a composite end point of death, placement of LV assist device, or heart transplant after CRT implant. In the testing cohort, we compared the discriminative ability of the ML Response Score and current guidelines to predict risk of reaching the composite end point. We also compared event-free survival in the subgroups from ML Response Score quartiles.

Stratification and Reclassification of Current Guidelines

Our final goal was to assess how the ML Response Score could add value to current guidelines. In the testing cohort, we calculated the mean and SD ML Response Score for 3 groups defined by class I, IIa, and IIb criteria, respectively. Within each group, we compared echocardiographic response rates and event-free survival in patients with scores below and above the respective group mean. We then reclassified patients based on ML Response Score. Class I patients with scores below the class I mean–SD were reclassified to low class I. Class IIa patients with scores above the class IIa mean+SD were reclassified to high class IIa, and patients with scores below the class IIa mean–SD were reclassified to low class IIa. Class IIb patients with scores above the class IIb mean–SD were reclassified to high class IIb. Echocardiographic and survival outcomes were compared in low class I versus high class IIa and in low class IIa versus high class IIb.

Statistical Analysis

During classifier development, mean AUC comparisons were assessed with a paired sample *t* test modified for multiple iterations of 5-fold cross-validation.[21] During the testing phase, single AUC comparisons were computed with the nonparametric approach by DeLong et al.[22] At the optimal operating point, McNemar test was used to assess classification differences compared with class I criteria.[23] Descriptive statistics between ML subgroups were assessed with χ^2 test on categorical variables, 1-way ANOVA on continuous normally distributed variables, and the Kruskal-Wallis test for continuous non-normally distributed variables.

Kaplan-Meier curves were used to visualize survival. Cox proportional hazards models were used to compute hazard ratios between subgroups. Harrell concordance index was used to quantify the discriminative ability for prediction of time-to-event data with censoring.[24] Significance of concordance index differences was assessed using the nonparametric approach by Kang et al.[25] Two-sided $P<0.05$ assessed significance. In comparing baseline characteristics among ML quartiles, Bonferroni correction was used to assess significance.

Software

ML analyses, survival curve visualization, Cox proportional hazards models, and k-fold cross-validation paired sample *t* test were conducted in MATLAB 2018b (MathWorks, Natick, MA). Minimum redundancy and maximum relevance feature selection was conducted using the Feature Selection Library MATLAB toolbox.[26] Construction of the online calculator (Methods in the Data Supplement) and all remaining statistical analyses were computed in RStudio 1.1.456 (RStudio, Boston, MA) with R 3.5.1 (R Foundation, Vienna, Austria) with packages pROC,[27] compareC,[25] e1071, and tableone.

RESULTS

One thousand four hundred ninety-five Cleveland Clinic patients and 253 Johns Hopkins patients underwent CRT implantation during the reviewed period. Nine hundred thirty-two patients from Cleveland Clinic and 251 patients from Johns Hopkins had pre- and post-CRT LVEF measurements. Of these, 690 from Cleveland Clinic and 235 from Johns Hopkins had complete data for analysis. The training cohort consisted of 470 patients (235, Cleveland Clinic; 235, Johns Hopkins). The testing cohort consisted of 455 Cleveland Clinic patients. Positive CRT response criteria were met by 290 (42%) patients from Cleveland Clinic, 95 (40%) patients from Johns Hopkins, 196 (42%) patients from the training cohort, and 189 (42%) patients from the testing cohort. LVEF improvement ≥5% was experienced by 427 (62%) patients from Cleveland Clinic, 134 patients from Johns Hopkins (57%), 285 (61%) patients from the training cohort, and 276 (61%) patients from the testing cohort. Baseline characteristics are provided in Table 1.

Classifier Development: Cross-Validation and Feature Selection

Performance of the feature set and ML algorithm combinations during classifier development are provided in Table 2. Guidelines had a mean AUC of 0.64. The highest performing classifiers used the naive Bayes algorithm with the minimal feature set and had better response prediction than guidelines (mean AUC, 0.72; $P<0.001$). A learning curve is provided in Figure III in the Data Supplement. Adding physical characteristics, comorbidities, and pharmacotherapy did not improve response prediction. None of the feature selection algorithms improved

Table 1. Baseline Characteristics

Variable	Entire Cohort (n=925)	Cleveland Clinic (n=690)	Johns Hopkins (n=235)	Training Cohort (n=470)	Testing Cohort (n=455)
Physical characteristics					
Age, y	65.6±12.6	65.7±12.5	64.7±13.0	65.0±12.6	65.9±12.6
Male sex	605 (65%)	447 (65%)	158 (67%)	304 (65%)	301 (66%)
Weight, kg	84.5±21.7	85.2±21.1	84.3±23.5	84.5±22.2	85.4±21.2
Height, inches	67.7±4.1	67.7±4.0	67.9±4.3	67.7±4.2	67.7±4.0
Body mass index, kg/m^2	28.6±6.4	28.7±6.0	28.3±7.3	28.5±6.7	28.8±6.0
Heart failure, echocardiography, and ECG characteristics					
NYHA class					
I	13 (1%)	7 (1%)	6 (3%)	9 (2%)	4 (1%)
II	131 (14%)	80 (12%)	51 (22%)	83 (18%)	48 (11%)
III	749 (81%)	574 (83%)	175 (75%)	366 (78%)	383 (84%)
IV	32 (4%)	29 (4%)	3 (1%)	12 (3%)	20 (4%)
LVEF, %	23.1±9.1	23.8±9.4	21.2±7.9	22.7±9.1	23.6±9.1
LV end-diastolic diameter, cm	6.1±1.1	6.0±1.1	6.3 ±1.1	6.1±1.1	6.0±1.1
Ischemic cardiomyopathy	465 (50%)	344 (50%)	121 (51%)	231 (49%)	234 (51%)
QRSd, ms	158±29	156±29	165±26	160±29	156±29
LBBB	460 (50%)	325 (47%)	135 (57%)	237 (50%)	223 (49%)
Right ventricular pacing	163 (18%)	109 (16%)	54 (23%)	92 (20%)	71 (16%)
Right bundle branch block	77 (8%)	60 (9%)	17 (7%)	41 (9%)	36 (8%)
Nonspecific intraventricular conduction delay	162 (18%)	133 (19%)	29 (12%)	72 (15%)	90 (20%)
Atrial fibrillation	509 (55%)	360 (52%)	149 (63%)	271 (58%)	238 (52%)
Comorbidity characteristics					
Hypertension	611 (66%)	435 (63%)	176 (75%)	324 (69%)	287 (63%)
Hyperlipidemia	573 (62%)	424 (61%)	149 (63%)	293 (62%)	280 (62%)
Chronic obstructive pulmonary disease	169 (18%)	113 (16%)	56 (24%)	98 (21%)	71 (16%)
Cerebrovascular accident or transient ischemic attack	124 (13%)	71 (10%)	53 (23%)	80 (17%)	44 (10%)
Malignancy	133 (14%)	99 (14%)	34 (14%)	67 (14%)	66 (15%)
Diabetes mellitus	346 (37%)	251 (36%)	95 (40%)	172 (37%)	174 (38%)
Creatinine, mg/dL	1.1 (0.9–1.5)	1.1 (0.9–1.5)	1.2 (1.0–1.5)	1.1 (0.9–1.5)	1.1 (0.9–1.5)
White blood cell count (10^9 cells/L)	7.4±2.5	7.4±2.5	7.5±2.5	7.5±2.6	7.4±2.5
Hemoglobin, g/dL	12.7±1.9	12.7±1.9	12.7±1.9	12.8±1.8	12.7±1.9
Red blood cell distribution width	14.5 (13.5–15.9)	14.4 (13.4–15.7)	14.9 (13.8–16.3)	14.6 (13.5–16.0)	14.5 (13.5–15.7)
Coronary artery bypass graft	306 (33%)	234 (34%)	72 (31%)	156 (33%)	150 (33%)
Percutaneous coronary intervention	193 (21%)	139 (20%)	54 (23%)	94 (20%)	99 (22%)
Atrioventricular node ablation	85 (9%)	46 (7%)	39 (17%)	57 (12%)	28 (6%)
Pharmacotherapy characteristics					
β-Blocker	796 (86%)	589 (85%)	207 (88%)	413 (88%)	383 (84%)
ACE inhibitor or ARB	762 (82%)	560 (81%)	202 (86%)	405 (86%)	357 (78%)
Diuretic	748 (81%)	539 (78%)	210 (89%)	397 (84%)	352 (77%)
Nitrate	174 (19%)	152 (22%)	22 (9%)	69 (15%)	105 (23%)
Hydralazine	89 (10%)	70 (10%)	19 (8%)	40 (9%)	49 (11%)
Aldosterone antagonist	306 (33%)	228 (33%)	78 (33%)	161 (34%)	145 (32%)
Statin	534 (58%)	385 (56%)	149 (63%)	279 (59%)	255 (56%)

(*Continued*)

Table 1. Continued

Variable	Entire Cohort (n=925)	Cleveland Clinic (n=690)	Johns Hopkins (n=235)	Training Cohort (n=470)	Testing Cohort (n=455)
Warfarin	331 (36%)	216 (31%)	115 (49%)	191 (41%)	140 (31%)
Clopidogrel	128 (14%)	100 (14%)	28 (12%)	60 (13%)	68 (15%)
Digoxin	302 (33%)	214 (31%)	88 (37%)	153 (33%)	149 (33%)
Antiarrhythmic drugs	159 (17%)	107 (16%)	52 (22%)	86 (18%)	73 (16%)
Procedural characteristics					
Epicardial LV lead	80 (9%)	74 (11%)	6 (3%)	34 (7%)	46 (10%)

Continuous normally distributed variables are reported as mean±SD. Continuous non-normally distributed variables are reported as median (interquartile range). Categorical variables are reported as n (%). ACE indicates angiotensin-converting enzyme; ARB, angiotensin receptor blocker; LBBB, left bundle branch block; LV, left ventricle; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; and QRSd, QRS duration.

performance beyond the minimal feature set. Feature selection algorithm rankings are provided in Figures IV through VII in the Data Supplement . Logistic regression performed comparably (mean AUC, 0.71), and model details are provided in Table I in the Data Supplement.

Classifier Testing

ROC curves for echocardiographic CRT response prediction on the testing data are shown in Figure 2. The ML classifier had significantly better prediction than guidelines (AUC, 0.70 versus 0.65; *P*=0.012) and had comparable performance to the trained logistic regression model (AUC, 0.72; Table I in the Data Supplement). The ML classifier did not have institution-specific performance variation (Figure VIII in the Data Supplement). The ROC curve from the modified CRT response score had lower predictive performance than guidelines (AUC, 0.61 versus 0.65; *P*=0.10). Remaining predictive metrics are shown in Table 3. Of current guidelines, class I criteria had the most accurate prediction (accuracy, 0.63; sensitivity, 0.46; specificity, 0.76; positive predictive value, 0.57; negative predictive value, 0.66). At the specified operating point, the ML classifier had significantly improved prediction (accuracy, 0.68; sensitivity, 0.46; specificity, 0.83; positive predictive value, 0.67; negative predictive value, 0.69; *P*=0.02).

The ML Response Score was also predictive of other echocardiographic outcomes. Compared with current guidelines, the ML Response Score trended toward improved prediction of response defined by LVEF improvement ≥5% (AUC, 0.68 versus 0.64; *P*=0.065; Figure IX in the Data Supplement) and significantly improved prediction of super response (AUC, 0.76 versus 0.70; *P*=0.005; Figure X in the Data Supplement). ML predictions were sensitive but not specific among class I patients, were less sensitive but more specific among class IIa patients, and generally predicted nonresponse among class IIb patients (Table II in the Data Supplement).

Descriptive characteristics of the ML Response Score quartiles are provided in Table III in the Data Supplement. Among significant variables after Bonferroni correction (*P*<0.00125), quartiles predicting favorable CRT response had lower proportion of men, shorter height, lower LVEF, lower proportion of ischemic cardiomyopathy, longer QRSd, higher proportion of LBBB, lower proportion of atrial fibrillation, lower creatinine, lower proportion of coronary artery bypass graft, and lower proportion of nitrate and statin use.

Survival Analysis

Median follow-up was 6.9 years (interquartile range, 3.9–10.0 years). Two hundred forty-three patients (53%)

Table 2. Comparison of Machine Learning Classifier Area Under the Curve During Cross-Validation

	Feature Selection Method						
Classifier Algorithm	Minimal	Comorbidities	Broad	mRMR	Random Forest	Univariate	Lasso
Logistic regression	0.71±0.05*	0.69±0.05†	0.68±0.05†	0.67±0.05	0.68±0.05†	0.68±0.05†	0.69±0.05†
Diagonal linear discriminant analysis	0.70±0.05*	0.68±0.05†	0.67±0.05†	0.66±0.05	0.67±0.05†	0.67±0.05	0.67±0.05
Naive Bayes	0.72±0.05‡§	0.70±0.05*	0.70±0.05†	0.67±0.05	0.69±0.05†	0.68±0.05†	0.68±0.05†
Support vector machine	0.70±0.05*	0.69±0.05†	0.67±0.05	0.66±0.05	0.68±0.05†	0.67±0.05†	0.68±0.05†
Random forest	0.71±0.05*	0.69±0.05†	0.69±0.05†	0.66±0.05	0.68±0.05†	0.68±0.05†	0.69±0.05†
AdaBoost	0.64±0.05	0.63±0.05	0.63±0.05	0.60±0.05	0.62±0.05	0.62±0.05	0.62±0.05

Reported as mean±SD. mRMR indicates minimal redundancy maximum relevance.
*P<0.005, †P<0.05, ‡P<0.0005 compared with current guidelines.
§Highest-performing feature set-learning algorithm combination.

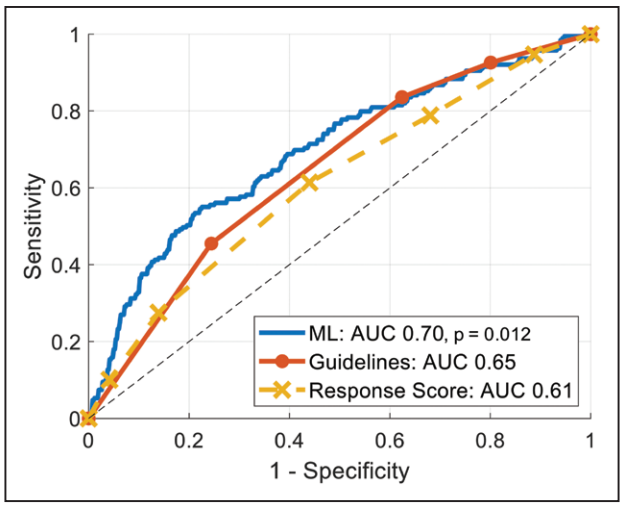


Figure 2. Prediction of echocardiographic response to cardiac resynchronization therapy (CRT): receiver operating characteristic (ROC) curves. Area under the curve (AUC) was computed for 3 different models on the testing cohort. Machine learning (ML) response score is generated by the naive Bayes classifier and minimal feature set. Guidelines: ROC constructed from class I, IIa, and IIb guidelines. Response score: modified CRT response score using female sex, nonischemic cardiomyopathy, left bundle branch block, QRS duration, and left ventricular end-diastolic diameter[7] (Methods in the Data Supplement).

experienced the composite end point within 10-year follow-up. ML Response Scores showed greater discrimination of risk for the composite end point than current guidelines (concordance index, 0.61 versus 0.56; *P*<0.001). ML Response Score quartiles had significantly different event-free survival (Figure 3). Patients in the fourth quartile were most likely to reach the composite end point, whereas patients in the first quartile were least likely (hazard ratio, 0.34; *P*<0.001). Current guidelines appropriately stratified event-free survival (hazard ratio, 0.48; *P*<0.001, between class I and IIb; Figure 4).

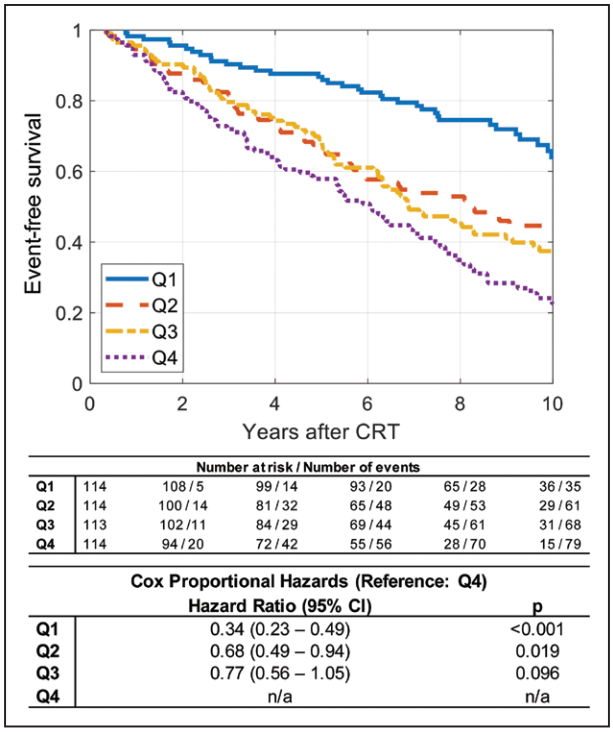
Stratification and Reclassification of Current Guidelines

The mean ML Response Score was 0.67±0.17 in class I patients, 0.40±0.23 in class IIa patients, and 0.15±0.11 in class IIb patients. Class I patients with above average ML Response Scores above the class I mean had greater response rates and reduced risk for the composite end point (Figure 5), while stratification of class IIa and IIb patients was less pronounced (Figures XI and XII in the Data Supplement). Based on ML Response Scores, 23 (15%) class I patients were reclassified to low class I and 36 (21%) class IIa patients were reclassified to high class IIa. The high class IIa group had improved echocardiographic outcomes and greater long-term survival than low class I (Figure 6). Thirty-one (18%) class IIa patients were reclassified to low class IIa, and 8 (13%) class IIb patients were reclassified to high class IIb. The high class IIb group had improved long-term survival

Table 3. Response Prediction Performance of Clinical Predictors and ML on the Testing Cohort

Predictor	Accuracy	Sensitivity	Specificity	PPV	NPV	*P* Value
Response definition: LVEF improvement ≥10%						
Class I criteria	0.63	0.46	0.76	0.57	0.66	NA
Class I or IIa criteria	0.57	0.84	0.38	0.49	0.76	0.028
Class I, IIa, or IIb criteria	0.50	0.93	0.20	0.45	0.79	<0.001
QRSd ≥150 ms	0.58	0.74	0.47	0.50	0.71	0.043
LBBB	0.59	0.60	0.59	0.51	0.67	0.034
QRSd ≥150 ms and LBBB	0.62	0.48	0.72	0.55	0.66	0.30
Nonischemic cardiomyopathy	0.57	0.57	0.58	0.49	0.65	0.065
Female sex	0.58	0.41	0.71	0.50	0.63	0.12
ML: naive Bayes classifier with the minimal feature set	0.68	0.46	0.83	0.67	0.69	0.018
Response definition: LVEF improvement ≥5%						
Class I criteria	0.56	0.41	0.79	0.75	0.47	NA
Class I or IIa criteria	0.64	0.79	0.41	0.68	0.56	0.005
Class I, IIa, or IIb criteria	0.63	0.90	0.22	0.64	0.58	0.044
QRSd ≥150 ms	0.61	0.69	0.50	0.68	0.51	0.043
LBBB	0.58	0.56	0.62	0.70	0.48	0.24
QRSd ≥150 ms and LBBB	0.56	0.44	0.75	0.73	0.47	0.30
Nonischemic cardiomyopathy	0.56	0.54	0.60	0.67	0.46	0.50
Female sex	0.52	0.38	0.73	0.69	0.44	0.16
ML: naive Bayes classifier with the minimal feature set	0.67	0.73	0.58	0.73	0.58	<0.001

Predictions compared with class I criteria. LBBB indicates left bundle branch block; LVEF, left ventricular ejection fraction; ML, machine learning; NPV, negative predictive value; PPV, positive predictive value; and QRSd, QRS duration.

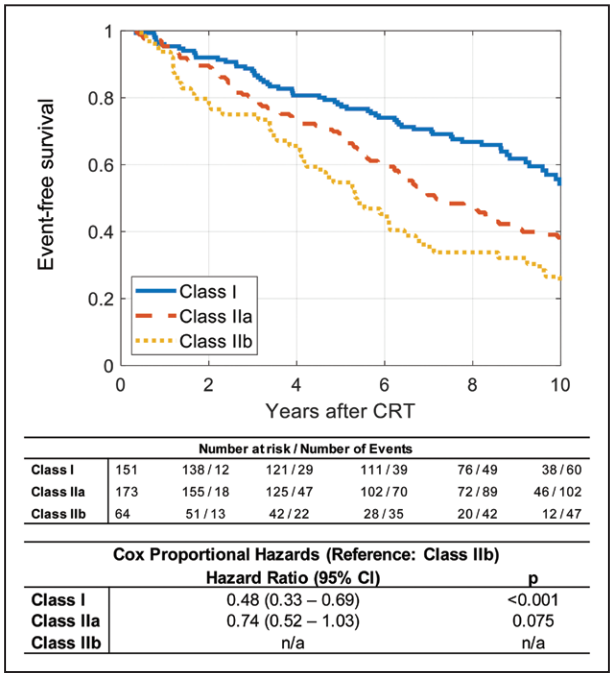


Number at risk / Number of events						
Q1	114	108 / 5	99 / 14	93 / 20	65 / 28	36 / 35
Q2	114	100 / 14	81 / 32	65 / 48	49 / 53	29 / 61
Q3	113	102 / 11	84 / 29	69 / 44	45 / 61	31 / 68
Q4	114	94 / 20	72 / 42	55 / 56	28 / 70	15 / 79

Cox Proportional Hazards (Reference: Q4)		
	Hazard Ratio (95% CI)	p
Q1	0.34 (0.23 – 0.49)	<0.001
Q2	0.68 (0.49 – 0.94)	0.019
Q3	0.77 (0.56 – 1.05)	0.096
Q4	n/a	n/a

Figure 3. Event-free survival after cardiac resynchronization therapy (CRT) implant across machine learning response score quartiles.
Kaplan-Meier curves for quartiles of the machine learning response score. The outcome was a composite of death, heart transplantation, and placement of left ventricular assist device.

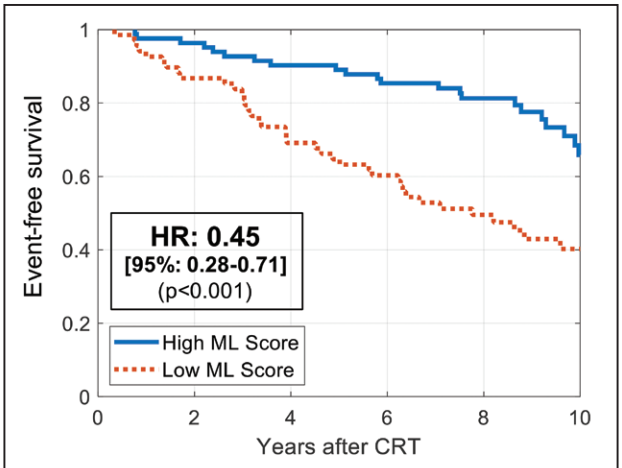
compared with low class IIa, but echocardiographic response rates were equal (Figure XIII in the Data Supplement). Response rates in subgroups defined by ML



Number at risk / Number of Events						
Class I	151	138 / 12	121 / 29	111 / 39	76 / 49	38 / 60
Class IIa	173	155 / 18	125 / 47	102 / 70	72 / 89	46 / 102
Class IIb	64	51 / 13	42 / 22	28 / 35	20 / 42	12 / 47

Cox Proportional Hazards (Reference: Class IIb)		
	Hazard Ratio (95% CI)	p
Class I	0.48 (0.33 – 0.69)	<0.001
Class IIa	0.74 (0.52 – 1.03)	0.075
Class IIb	n/a	n/a

Figure 4. Event-free survival after cardiac resynchronization therapy (CRT) implant across current guidelines.
Kaplan-Meier curves for patients meeting different guideline criteria. The outcome was a composite of death, heart transplantation, and placement of left ventricular assist device.



Number at risk					
High ML (n = 83)	79	74	70	51	24
Low ML (n = 68)	59	47	41	25	14

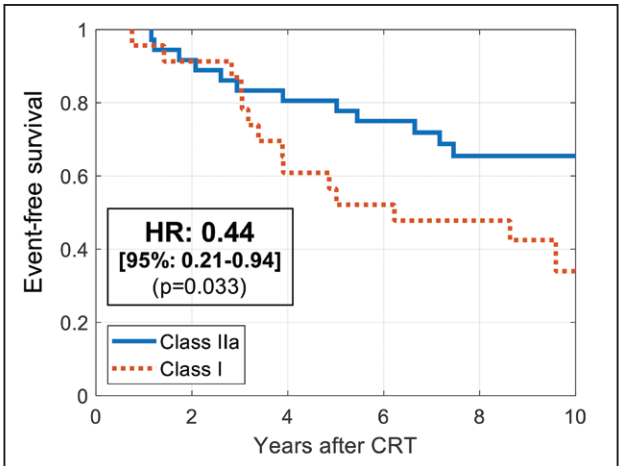
Echocardiographic Response Rates			
Threshold	Low ML	High ML	p
LVEF ≥ +5%	46 (68%)	68 (82%)	0.042
LVEF ≥ +10%	26 (38%)	60 (72%)	<0.001
LVEF ≥ +20%	12 (18%)	35 (42%)	0.001

Figure 5. Machine learning (ML) response score stratification of cardiac resynchronization therapy (CRT) patients with class I recommendations.
Class I patients with above average ML Response Scores comprised the high-score group, and patients with below average scores comprised the low-score group. Kaplan-Meier curves depict survival free of a composite of death, heart transplantation, and placement of left ventricular assist device. Echocardiographic response rates are provided. HR indicates hazard ratio; and LVEF, left ventricular ejection fraction.

Response Score levels and current guideline criteria (Table IV in the Data Supplement) were used for the online calculator.

DISCUSSION

In this study, we evaluated ML prediction of CRT response using common clinical variables with 3 following principal findings: (1) a trained ML classifier predicted echocardiographic CRT response with higher performance than all clinical predictors, including LBBB, QRSd ≥150 ms, and class I guideline criteria; (2) the ML prediction for echocardiographic response was also predictive of long-term survival; (3) the optimal classifier was developed from a minimal feature set that only used 9 readily available clinical variables: sex, QRSd, QRS morphology (LBBB versus right bundle branch block versus intraventricular conduction delay versus right ventricular pacing), cause of cardiomyopathy (ischemic or nonischemic), LVEF, LV end-diastolic diameter, NYHA classification, history



Number at risk					
High Class IIa (n = 36)	33	29	27	19	16
Low Class I (n = 23)	21	14	12	9	4

Echocardiographic Response Rates

Threshold	Low Class I	High Class IIa	p
LVEF ≥ +5%	14 (61%)	27 (75%)	0.25
LVEF ≥ +10%	6 (26%)	21 (58%)	0.015
LVEF ≥ +20%	1 (4%)	11 (31%)	0.015

Figure 6. Machine learning (ML) response score reclassification of cardiac resynchronization therapy (CRT) patients with class I and IIa recommendations.
Class IIa patients with high ML Response Scores were reclassified to the high class IIa group, and class I patients with low scores were reclassified to the low class I group. Kaplan-Meier curves depict survival free of a composite of death, heart transplantation, and placement of left ventricular assist device. Echocardiographic response rates are provided. HR indicates hazard ratio; and LVEF, left ventricular ejection fraction.

of atrial fibrillation, and the presence of epicardial LV lead. Adding other comorbidities, pharmacologic information, and feature selection algorithms did not improve prediction.

Appropriate CRT patient selection is an important goal, and our study demonstrated that ML can leverage clinical data to improve CRT outcome prediction. A common pitfall in ML models is overfitting data to outcomes. However, the reliability of our model was confirmed by several findings. Classification performance was comparable in the training and testing cohorts across 2 institutions, and the output of our ML model predicted multiple levels of echocardiographic response (LVEF improvement ≥5%, ≥10%, and ≥20%) and discriminated long-term survival. With respect to current guidelines, the ML Response Score stratified outcomes within class I patients and identified a portion of class IIa patients with better outcomes than a portion of class I patients. The ML Response Score did not convincingly identify class IIb patients with favorable outcomes.

The feature selection in our study design assisted in interpreting ML output. Incorporating noncontributory features can lead to overfitting, identifying false relationships between features and outcomes. A balance between insufficient and excessive inclusion of features is important for optimizing ML classification.[11] We evaluated 7 feature sets, improving our understanding of which features are important in obtaining accurate CRT response prediction. Feature selection algorithms failed to improve classifier performance beyond the minimal feature set, with creatinine, history of coronary artery bypass graft, and red blood cell distribution width being the only additional variables that were identified as a top-15 feature by at least 3 of 4 feature selection algorithms. Adding comorbidity and pharmacotherapy variables to the minimal feature set also reduced classification performance, likely because of model overfitting.

Our results conflict with findings by Zeitler et al[9] identifying comorbidities as important predictors of CRT response. There are several potential explanations. Zeitler et al[9] analyzed MADIT-CRT (Multicenter Automatic Defibrillator Implantation Trial With Cardiac Resynchronization Therapy) NYHA I/II patients with LBBB, while we had diverse representation of conduction morphology and limited NYHA I/II patients. They included comorbidities that were not available in our data, such as history of ventricular arrhythmias and current smoking. Additionally, they included coronary artery disease as a comorbidity, which was represented by ischemic cardiomyopathy in our model. And interestingly, although our optimal ML classifier did not incorporate physical characteristics, comorbidities, or pharmacotherapy variables, significant differences still existed among some of these variables between the ML Response Score quartiles. Favorable quartiles had significantly lower creatinine levels, less history of coronary artery bypass graft, and lower nitrate, statin, and antiarrhythmic usage. This suggests an interdependence between these variables and the 9 variables included in the minimal ML classifier.

The optimal learning algorithm may provide insight into the relationship between CRT response predictors and outcomes, as their respective prediction performance depends on how features are related to classifications.[11] Logistic regression quantifies the effect of features on classification odds. Linear discriminant analysis uses linear combinations of features to separate classes. Support vector machines identify hyperplanes in high-dimensional space to separate classes. Naive Bayes classifiers use conditional probabilities with naive interfeature independence assumptions. Random forests use a large ensemble of weakly predictive decision trees to develop a single stronger classifier. In our study, a naive Bayes classifier had the highest performance during cross-validation. However,

logistic regression trained with the minimal feature set had nearly equal performance during cross-validation (AUC, 0.71 versus 0.72) and slightly better performance when evaluated on the testing set post hoc (AUC, 0.72 versus 0.70). Linear models showing comparable performance to nonlinear algorithms support the notion that CRT response prediction via clinical variables is largely driven by simple relationships with relatively few variables.

Our ML study design suggests that improving CRT response prediction does not require more advanced methods to discover abstract relationships between commonly available clinical variables and CRT response. Although ML significantly improved prediction compared with guidelines, it is important to note that the prediction improvement was marginal, with AUC improvements of 0.05 to 0.08. When ML classifiers do not perform at a high level, it may suggest that features are not sufficiently discriminative. Rather, new features that are more predictive of CRT response should be further investigated. Another possible explanation for limited predictive performance is the size of the training set. Our learning curve suggests that predictive performance has nearly but not completely plateaued at our training set size. Predictive performance may increase with even larger training sets, as this may help nonlinear models capture interactions between variables.

Strengths

We developed an ML model to predict echocardiographic CRT response and discriminate long-term survival using a large set of observational data from 2 cohorts and an independent validation set, reinforcing the generalizability of the model.[28] The model showed minimal institution-specific performance variation. These results validate previous ML studies in CRT.[12,13] We compared ML performance to common clinical predictors and current practice recommendations. Although ML provided the best predictions, our data also support the utility of simpler decision-making via current guidelines. Our ML evaluation was rigorous, evaluating many different classifiers and feature selection methods. We provided interpretability, which may help guide future studies investigating ML and CRT response prediction.

Limitations

All patients in the study received CRT; thus we did not assess benefit compared with a non-CRT comparison group. Echocardiogram follow-up durations were variable. Many variables were coded in a binary fashion, such as history of atrial fibrillation or ischemic cardiomyopathy, resulting in incomplete representation of the spectrum of these conditions. Although we used data from 2 centers, we only had survival data for the Cleveland Clinic cohort. It remains unknown how a model with even larger training set size may perform. There are additional preimplant features predictive of CRT response, such as frailty index,[29] PR interval,[30] echocardiogram metrics,[31] and magnetic resonance measurements,[32] which were not incorporated into our model. In addition to preimplant variables, procedural features such as lead location[33] may be incorporated into future models useful for counseling patients regarding expected outcomes after the procedure. Our study population mainly met class I, IIa, or IIb criteria. An exciting direction of future ML studies is to identify patients outside of these criteria who may benefit from CRT.

Conclusions

An ML classifier incorporating 9 common clinical variables improved prediction of echocardiographic CRT response and survival when compared with current guidelines. Models incorporating greater amounts of clinical variables did not improve performance. Significant remaining limitations confirm a need to better identify new variables in CRT response prediction. We have made a calculator for the ML Response Score available at http://riskcalc.org:3838/CRTResponseScore to assist in shared decision-making.

ARTICLE INFORMATION

Received February 18, 2019; accepted April 23, 2019.

Guest Editor for this article was Andrew E. Epstein, MD.

The Data Supplement is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCEP.119.007316.

Correspondence

Mina K. Chung, MD, Department of Cardiovascular Medicine, Heart and Vascular Institute, Cleveland Clinic, 9500 Euclid Ave, J2-2, Cleveland, OH 44195. Email chungm@ccf.org

Affiliations

Cleveland Clinic Lerner College of Medicine (A.K.F., M.A.L.) and Department of Biomedical Engineering (A.M.), Case Western Reserve University, OH. Department of Cardiovascular Medicine, Heart and Vascular Institute (J.R., D.P., S.T., K.M.T., N.V., M.J.N., E.Z.G., R.A.G., M.K.C.) and Department of Quantitative Health Sciences, Lerner Research Institute (X.J., J.B.), Cleveland Clinic, OH. Division of Cardiology, The Johns Hopkins Hospital, Baltimore, MD (C.J.P., S.S., D.D.S.). Louis Stokes Cleveland Veterans Administration Medical Center, OH (A.M.).

Sources of Funding

This study was supported by National Institutes of Health (NIH)/National Heart, Lung, and Blood Institute grant R01-HL111314; an American Heart Association Atrial Fibrillation Strategically Focused Research Network grant; NIH UL1-RR024989—NIH National Center for Research Resources for Case Western Reserve University and Cleveland Clinic Clinical and Translational Science Award; and by the Center of Excellence in Cardiovascular Translational Functional Genomics, Heart and Vascular Institute and Lerner Research Institute funds. This study was also supported by Tomsich Atrial Fibrillation Research Fund; Heart and Vascular Institute and Lerner Research Institute Philanthropy funds; National Cancer Institute/NIH: 1U24CA199374-01, R01CA202752-01A1, R01CA208236-01A1, R01-CA216579-01A1, R01-CA220581-01A1, 1U01-CA239055-01; National Center for Research Resources 1-C06 RR12463-01;

received Veterans Affairs (VA) Merit Review Award IBX004121A from the VA Biomedical Laboratory Research and Development Service; Department of Defense (DOD) Prostate Cancer Idea Development Award (W81XWH-15-1-0558); DOD Lung Cancer Investigator-Initiated Translational Research Award (W81XWH-18-1-0440). This study was also supported by DOD Peer-Reviewed Cancer Research Program (W81XWH-16-1-0329); Ohio Third Frontier Technology Validation Fund; and Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering and the Clinical and Translational Science Award Program at Case Western Reserve University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, VA, DOD, or US Government.

Disclosures

Dr Rickard has been a speaker for Boston Scientific and consultant for Medtronic. Dr Niebauer has received research support from Biosense Webster, Biotronik, Boston Scientific, Medtronic, St. Jude Medical, and Sorin. Dr Varma has received consulting fees and honoraria from St. Jude Medical, Boston Scientific, Sorin, Biotronik, and Medtronic. Dr Gorodeski has received research support and consulting fees from Abott. Dr Madabhushi is an equity holder in Elucid Bioimaging and in Inspirata. He is also a scientific advisory consultant for Inspirata. In addition, he has served as a scientific advisory board member for Inspirata, Astrazeneca, and Merck. He also has sponsored research agreements with Philips and Inspirata. His technology has been licensed to Elucid Bioimaging and Inspirata. He is also involved in a NIH U24 grant with PathCore, and 3 different R01 grants with Inspirata. Dr Chung serves on a steering committee and is a speaker for EPIC Alliance—a forum for networking and mentoring of women in cardiac electrophysiology sponsored by Biotronik but declines all honoraria from device companies. The other authors report no conflicts.

REFERENCES

1. Abraham WT, Fisher WG, Smith AL, Delurgio DB, Leon AR, Loh E, Kocovic DZ, Packer M, Clavell AL, Hayes DL, Ellestad M, Trupp RJ, Underwood J, Pickering F, Truex C, McAtee P, Messenger J; MIRACLE Study Group. Multicenter InSync Randomized Clinical Evaluation. Cardiac resynchronization in chronic heart failure. *N Engl J Med.* 2002;346:1845–1853. doi: 10.1056/NEJMoa013168
2. Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De Marco T, Carson P, DiCarlo L, DeMets D, White BG, DeVries DW, Feldman AM; Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) Investigators. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N Engl J Med.* 2004;350:2140–2150. doi: 10.1056/NEJMoa032423
3. Cleland JG, Daubert JC, Erdmann E, Freemantle N, Gras D, Kappenberger L, Tavazzi L; Cardiac Resynchronization-Heart Failure (CARE-HF) Study Investigators. The effect of cardiac resynchronization on morbidity and mortality in heart failure. *N Engl J Med.* 2005;352:1539–1549. doi: 10.1056/NEJMoa050496
4. Vernooy K, van Deursen CJ, Strik M, Prinzen FW. Strategies to improve cardiac resynchronization therapy. *Nat Rev Cardiol.* 2014;11:481–493. doi: 10.1038/nrcardio.2014.67
5. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJ, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WH, Tsai EJ, Wilkoff BL; American College of Cardiology Foundation; American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2013;62:e147–e239. doi: 10.1016/j.jacc.2013.05.019
6. Varma N, Manne M, Nguyen D, He J, Niebauer M, Tchou P. Probability and magnitude of response to cardiac resynchronization therapy according to QRS duration and gender in nonischemic cardiomyopathy and LBBB. *Heart Rhythm.* 2014;11:1139–1147. doi: 10.1016/j.hrthm.2014.04.001
7. Goldenberg I, Moss AJ, Hall WJ, Foster E, Goldberger JJ, Santucci P, Shinn T, Solomon S, Steinberg JS, Wilber D, Barsheshet A, McNitt S, Zareba W, Klein H; MADIT-CRT Executive Committee. Predictors of response to cardiac resynchronization therapy in the Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy (MADIT-CRT). *Circulation.* 2011;124:1527–1536. doi: 10.1161/CIRCULATIONAHA.110.014324
8. van Bommel RJ, Bax JJ, Abraham WT, Chung ES, Pires LA, Tavazzi L, Zimetbaum PJ, Gerritse B, Kristiansen N, Ghio S. Characteristics of heart failure patients associated with good and poor response to cardiac resynchronization therapy: a PROSPECT (Predictors of Response to CRT) sub-analysis. *Eur Heart J.* 2009;30:2470–2477. doi: 10.1093/eurheartj/ehp368
9. Zeitler EP, Friedman DJ, Daubert JP, Al-Khatib SM, Solomon SD, Biton Y, McNitt S, Zareba W, Moss AJ, Kutyifa V. Multiple comorbidities and response to cardiac resynchronization therapy: MADIT-CRT long-term follow-up. *J Am Coll Cardiol.* 2017;69:2369–2379. doi: 10.1016/j.jacc.2017.03.531
10. Gasparini M, Klersy C, Leclercq C, Lunati M, Landolina M, Auricchio A, Santini M, Boriani G, Proclemer A, Leyva F. Validation of a simple risk stratification tool for patients implanted with cardiac resynchronization therapy: the VALID-CRT risk score. *Eur J Heart Fail.* 2015;17:717–724.
11. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol.* 2017;69:2657–2664. doi: 10.1016/j.jacc.2017.03.571
12. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, Pouleur AC, Knappe D, Biering-Sørensen T, Kutyifa V, Moss A, Stein K, Solomon SD, Bijnens B. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail.* 2019;21:74–85. doi: 10.1002/ejhf.1333
13. Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, DeMets DL, Field ME, Eckhardt LL, Page CD. Machine learning algorithm predicts cardiac resynchronization therapy outcomes: lessons from the COMPANION Trial. *Circ Arrhythm Electrophysiol.* 2018;11:e005499. doi: 10.1161/CIRCEP.117.005499
14. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3:32–35.
15. Arshad A, Moss AJ, Foster E, Padeletti L, Barsheshet A, Goldenberg I, Greenberg H, Hall WJ, McNitt S, Zareba W, Solomon S, Steinberg JS; MADIT-CRT Executive Committee. Cardiac resynchronization therapy is more effective in women than in men: the MADIT-CRT (Multicenter Automatic Defibrillator Implantation Trial with Cardiac Resynchronization Therapy) trial. *J Am Coll Cardiol.* 2011;57:813–820. doi: 10.1016/j.jacc.2010.06.061
16. Rickard J, Michtalik H, Sharma R, Berger Z, Iyoha E, Green AR, Haq N, Robinson KA. Predictors of response to cardiac resynchronization therapy: a systematic review. *Int J Cardiol.* 2016;225:345–352. doi: 10.1016/j.ijcard.2016.09.078
17. Dupont M, Rickard J, Baranowski B, Varma N, Dresing T, Gabi A, Finucan M, Mullens W, Wilkoff BL, Tang WH. Differential response to cardiac resynchronization therapy and clinical outcomes according to QRS morphology and QRS duration. *J Am Coll Cardiol.* 2012;60:592–598. doi: 10.1016/j.jacc.2012.03.059
18. Garrigue S, Jaïs P, Espil G, Labeque JN, Hocini M, Shah DC, Haïssaguerre M, Clementy J. Comparison of chronic biventricular pacing between epicardial and endocardial left ventricular stimulation using Doppler tissue imaging in patients with heart failure. *Am J Cardiol.* 2001;88:858–862.
19. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell.* 2005;27:1226–1238. doi: 10.1109/TPAMI.2005.159
20. Haury AC, Gestraud P, Vert JP. The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *PLoS One.* 2011;6:e28210. doi: 10.1371/journal.pone.0028210
21. Bouckaert RR, Frank E. In: Dai H, Srikant R, Zhang C, eds. 8th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining Conference *Vol. LNAI 3056*. Berlin, Germany: Springer; 2004:3–12.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837–845.
23. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput.* 1998;10:1895–1923.
24. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* 1982;247:2543–2546.
25. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med.* 2015;34:685–703. doi: 10.1002/sim.6370
26. Roffo G. Feature selection library (MATLAB toolbox). arXiv:1607.01327 [cs] 2016. Available at: http://arxiv.org/abs/1607.01327.
27. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77. doi: 10.1186/1471-2105-12-77
28. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect

pneumonia in chest radiographs: a cross-sectional study. *PLoS Med.* 2018;15:e1002683. doi: 10.1371/journal.pmed.1002683
29. Kubala M, Guédon-Moreau L, Anselme F, Klug D, Bertaina G, Traullé S, Buiciuc O, Savouré A, Diouf M, Hermida JS. Utility of frailty assessment for elderly patients undergoing cardiac resynchronization therapy. *JACC Clin Electrophysiol.* 2017;3:1523–1533. doi: 10.1016/j.jacep.2017.06.012
30. Rickard J, Karim M, Baranowski B, Cantillon D, Spragg D, Tang WHW, Niebauer M, Grimm R, Trulock K, Wilkoff B, Varma N. Effect of PR interval prolongation on long-term outcomes in patients with left bundle branch block vs non-left bundle branch block morphologies undergoing cardiac resynchronization therapy. *Heart Rhythm.* 2017;14:1523–1528. doi: 10.1016/j.hrthm.2017.05.028
31. Tanaka H, Nesser HJ, Buck T, Oyenuga O, Jánosi RA, Winter S, Saba S, Gorcsan J III. Dyssynchrony by speckle-tracking echocardiography and response to cardiac resynchronization therapy: results of the Speckle Tracking and Resynchronization (STAR) study. *Eur Heart J.* 2010;31:1690–1700. doi: 10.1093/eurheartj/ehq213
32. Marsan NA, Westenberg JJ, Ypenburg C, van Bommel RJ, Roes S, Delgado V, Tops LF, van der Geest RJ, Boersma E, de Roos A, Schalij MJ, Bax JJ. Magnetic resonance imaging and response to cardiac resynchronization therapy: relative merits of left ventricular dyssynchrony and scar tissue. *Eur Heart J.* 2009;30:2360–2367. doi: 10.1093/eurheartj/ehp280
33. Saba S, Marek J, Schwartzman D, Jain S, Adelstein E, White P, Oyenuga OA, Onishi T, Soman P, Gorcsan J III. Echocardiography-guided left ventricular lead placement for cardiac resynchronization therapy: results of the speckle tracking assisted resynchronization therapy for electrode region trial. *Circ Heart Fail.* 2013;6:427–434. doi: 10.1161/CIRCHEARTFAILURE.112.000078